

From: Mehta, Ashwin
Sent: Tuesday, January 13, 2004 12:02 PM
T : STIC-Biotech/ChemLib
Subject: sequence search

STIC,

Please search the interference and commercial databases for the following from 09/991,262:

- 1) the nucleotide sequence of SEQ ID NO: 39
- 2) the amino acid sequences of SEQ ID NOs: 40 and 50

My mail room is REM 2C18, office 2A19, art unit 1638.

Thank you,
Ashwin

Ashwin Mehta
United States Patent and Trademark Office
Biotechnology Patent Examiner
(571) 272-0803

RECEIVED
JAN 13 2004
(611)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: _____
Date Completed: _____
Searcher Prep/Review: _____
Clerical: _____
Online time: _____

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: _____
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST (where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: _____
WWW/Internet: _____
Other (specify): _____

Gencore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2004, 21:39:51 ; Search time 12451 Seconds

(without alignments)
17453.372 Million cell updates/sec

Title: US-09-991-262-39

Perfect score: 5312

Sequence: 1 GTCTCTGCTCCCTCCCGACGCG.....CGAGCGCCCTGTGAGATACCA 5312

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank1:*

1: gb_ba:*

2: gb_hcg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_hcg_hum:*

31: em_hcg_inv:*

32: em_hcg_other:*

33: em_hcg_mus:*

34: em_hcg_pln:*

35: em_hcg_rod:*

36: em_hcg_mam:*

37: em_hcg_vtc:*

38: em_sy:*

39: em_htgo_hum:*

40: em_htgo_mus:*

41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5312	100.0	5312	6	AR125092
2	5312	100.0	5312	6	AR125093
3	5312	100.0	5312	6	AR125094
4	5281	99.4	5312	14	HAU18246
5	5206	98.0	5368	6	AR125095
6	88.4	1.7	6625	14	AF102884
7	76.4	1.4	6534	6	AR159699
8	76.4	1.4	6536	6	AX028972
9	76.4	1.4	6536	6	AX028972
10	71.2	1.3	2478	6	AR125096
11	71.2	1.3	2478	6	AR125097
12	71.2	1.3	2478	6	AR159698
13	71.2	1.3	2478	6	AX028950
14	71.2	1.3	2478	14	HYSCOAT
15	71.2	1.3	2478	14	AR125098
16	71.2	1.3	125020	9	AF429315
17	67.8	1.2	125020	9	AF429315
18	61.6	1.1	2000	6	AX655393
19	57.4	1.1	3865	1	ACVNFDPK
20	55	1.0	7203	1	ACVNF
21	55	1.0	2000	6	AX655393
22	54.6	1.0	157226	2	AP004840
23	53.6	1.0	202050	1	AL646065
24	52.6	1.0	8050	1	AE060633
25	51.8	1.0	302300	1	AP005034
26	51.6	1.0	298550	1	AP005047
27	51.2	1.0	115815	8	AP003874
28	51	1.0	126845	8	AP005244
29	51	1.0	146585	8	CNS08CAM
30	50.4	0.9	148449	2	CNS08C8R
31	50.4	0.9	1086	6	AX660360
32	50.2	0.9	12326	1	AE004990
33	49.8	0.9	11659	1	AE005131
34	49.8	0.9	299800	1	AP005040
35	49.8	0.9	2967	8	HVU507213
36	49.4	0.9	308050	1	SC0939124
37	49.4	0.9	3867	1	AY033407
38	48.8	0.9	295150	1	SC0939125
39	48.2	0.9	298450	1	SC0939107
40	48.2	0.9	203125	2	AC120986
41	47.8	0.9	300425	1	AP005041
42	47.8	0.9	1116	8	AR177966
43	47.6	0.9	1409	6	AF326486
44	47.6	0.9	301925	1	AP005046
45	47.6	0.9	301925	1	AP005046

ALIGNMENTS

RESULT 1

LOCUS AR125092

DEFINITION Sequence 39 from patent US 6177075.

ACCESSION AR125092

VERSION AR125092.1 GI:14111154

KEYWORDS

SOURCE

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 5312)

AUTHORS Christian,P.Daniel., Gordon,K.Henrich.Julius. and Hanzlik,T.Nelson.

TITLE Insect viruses and their uses in protecting plants

JOURNAL Patent: US 6177075-A 39 23-JAN-2001;

FEATURES Location/Qualifiers
 source 1..5312 /organism="unknown"
 BASE COUNT 1133 a 1740 c 1432 g 1007 t
 ORIGIN

Query Match 100.0%; Score 5312; DB 6; Length 5312;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

Oy 1 GTTCTGCTCCCGCGGACGTTAAATATAGGGGAACATGTACGGCAAGCGACAGACGCG 60
Db 1 GTTCTGCTCCCGCGGACGTTAAATATAGGGGAACATGTACGGCAAGCGACAGACGCG 60
Oy 61 GCGCGTGTCTACGCGCGGACAGATGTGCTTACCGGAAAGTACTGACAGAGACAGATC 120
Db 61 GCGCGTGTCTACGCGCGGACAGATGTGCTTACCGGAAAGTACTGACAGAGACAGATC 120
Oy 121 AAGTTGACCTTGCGCGCGGACCTGAAGGACCTAGAAACCTTCCACAGACTGTACTATCG 180
Db 121 AAGTTGACCTTGCGCGCGGACCTGAAGGACCTAGAAACCTTCCACAGACTGTACTATCG 180
Oy 181 CTGCGCTTCAAGGGGGGACCTTTACCCCGGACACAAACCCGATCTGGCGGGGACCA 240
Db 181 CTGCGCTTCAAGGGGGGACCTTTACCCCGGACACAAACCCGATCTGGCGGGGACCA 240
Oy 241 CGTGTGCGAAGAGAGATTCTGCACAATTTGCGCAGGGGACGTAGACAGATGCTCGAGATA 300
Db 241 CGTGTGCGAAGAGAGATTCTGCACAATTTGCGCAGGGGACGTAGACAGATGCTCGAGATA 300
Oy 301 GGGCGGTCTCTGCAACAGGCACTTAAAGTACATGGGGGACCGAAGCGCCCGTCCGACAC 360
Db 301 GGGCGGTCTCTGCAACAGGCACTTAAAGTACATGGGGGACCGAAGCGCCCGTCCGACAC 360
Oy 361 TATCAGGGGTGACCAAGTAGAGGGGACCCGCGAGGGGTGCGACACATTACGGCTTAGAG 420
Db 361 TATCAGGGGTGACCAAGTAGAGGGGACCCGCGAGGGGTGCGACACATTACGGCTTAGAG 420
Oy 421 TCTAGATCCGTGCGACAGGCGCGGCGGAGTTCAAGGCGGACGCTCACTGTGCGCAAC 480
Db 421 TCTAGATCCGTGCGACAGGCGCGGCGGAGTTCAAGGCGGACGCTCACTGTGCGCAAC 480
Oy 481 GGCATTGCTCCCGGACCTTCTGCGTCCAGCGAGTCCGCTCTTGGCGCTTCAAAATCGGCG 540
Db 481 GGCATTGCTCCCGGACCTTCTGCGTCCAGCGAGTCCGCTCTTGGCGCTTCAAAATCGGCG 540
Oy 541 GTTGAATTGGCAATCACTCCCTCTATAGCGTAGACCTTAGAGAGCTGGCCAAATCGGTT 600
Db 541 GTTGAATTGGCAATCACTCCCTCTATAGCGTAGACCTTAGAGAGCTGGCCAAATCGGTT 600
Oy 601 GAGAACCAAGCACTTCACATGTGTCGCGCGGTTCAATGACATGCGAGAAAGCTGCTTAC 660
Db 601 GAGAACCAAGCACTTCACATGTGTCGCGCGGTTCAATGACATGCGAGAAAGCTGCTTAC 660
Oy 661 ATGGAACAAGTGTAAATGCGAGCTCGGCTACCCCTTCCAGTATAGAAAGGCTATAG 720
Db 661 ATGGAACAAGTGTAAATGCGAGCTCGGCTACCCCTTCCAGTATAGAAAGGCTATAG 720
Oy 721 GGTGGAAGGACATGCGCATTTCCAGGGGGGGGACCTCCGTCACATTCCTTAGATTGAC 780
Db 721 GGTGGAAGGACATGCGCATTTCCAGGGGGGGGACCTCCGTCACATTCCTTAGATTGAC 780
Oy 781 TTTCATCAACGAGAGCCAGAGCGGCGCATCGAGAGGCTGCGCGCGGCTCTTAATCC 840
Db 781 TTTCATCAACGAGAGCCAGAGCGGCGCATCGAGAGGCTGCGCGCGGCTCTTAATCC 840
Oy 841 AGACCGCGCTCATTTTCTCGGCGACGACGACTGAGTGTGATGCTTACAGACTTC 900
Db 841 AGACCGCGCTCATTTTCTCGGCGACGACGACTGAGTGTGATGCTTACAGACTTC 900
Oy 901 CACAGATGGCTGCGCTACTACTGATGAGGAATACCCCACTCCGTTGGTTTCTCATCTC 960
Db 901 CACAGATGGCTGCGCTACTACTGATGAGGAATACCCCACTCCGTTGGTTTCTCATCTC 960

```

```

Oy 961 CATATAGAGTCCAGAGGCGCCACGCTCCAGCAATTAGACTGCGCATCATCTCGCGCCA 1020
Db 961 CATATAGAGTCCAGAGGCGCCACGCTCCAGCAATTAGACTGCGCATCATCTCGCGCCA 1020
Oy 1021 CCTGGAAGCGGACAGCTGCGCGGTGCTCCAGAGAGCTGCCAAGGCGCTGAGAAATCCA 1080
Db 1021 CCTGGAAGCGGACAGCTGCGCGGTGCTCCAGAGAGCTGCCAAGGCGCTGAGAAATCCA 1080
Oy 1081 AACATCTTTTATTAAGCGCGACGCTCGGGCACTAGAGCAATAGACATCTTACGTACAG 1140
Db 1081 AACATCTTTTATTAAGCGCGACGCTCGGGCACTAGAGCAATAGACATCTTACGTACAG 1140
Oy 1141 CACAAAGTCAACATGCTGCTCAATTTATAGCAAAACGCGTCTTGAAAGAACTAGTGAC 1200
Db 1141 CACAAAGTCAACATGCTGCTCAATTTATGCAAAACGCGTCTTGAAAGAACTAGTGAC 1200
Oy 1201 ATGACCGCTTGAATGCTTCCGCGCGGCTAGAGCTGGCGGGAATCGTGTGCTCAGAA 1260
Db 1201 ATGACCGCTTGAATGCTTCCGCGCGGCTAGAGCTGGCGGGAATCGTGTGCTCAGAA 1260
Oy 1261 GTCAACGAGAGCTCTGGAACATCTCACCGGCTGACCTGTCCGCACTGTGTCTT 1320
Db 1261 GTCAACGAGAGCTCTGGAACATCTCACCGGCTGACCTGTCCGCACTGTGTCTT 1320
Oy 1321 TACGCTCTCAATCATGACGCGCGGAAGGCTCGGCTGCTGTCAAGACCGCAAGAC 1380
Db 1321 TACGCTCTCAATCATGACGCGCGGAAGGCTCGGCTGCTGTCAAGACCGCAAGAC 1380
Oy 1381 GACGCTTTTGAAGAGACTTGTCTGAGAGTCTCAAGACAGTCTTGGGCTCTGTTGC 1440
Db 1381 GACGCTTTTGAAGAGACTTGTCTGAGAGTCTCAAGACAGTCTTGGGCTCTGTTGC 1440
Oy 1441 GGTCTGCGCAACTCAAAAGGACCGACGCTGCTTACTAAGCGCGTGTGATTAAGTAC 1500
Db 1441 GGTCTGCGCAACTCAAAAGGACCGACGCTGCTTACTAAGCGCGTGTGATTAAGTAC 1500
Oy 1501 CGAGTCACTGCTCGGAGACATATATGCGACGCTCGCTTACTAAGCGCGTGTGATTAAGTAC 1560
Db 1501 CGAGTCACTGCTCGGAGACATATATGCGACGCTCGCTTACTAAGCGCGTGTGATTAAGTAC 1560
Oy 1561 TTCCGCGGTCCCGGACCTGCGACCTGCGCGGCTTTACGACAGGAAAGACTTGAAGTTC 1620
Db 1561 TTCCGCGGTCCCGGACCTGCGACCTGCGCGGCTTTACGACAGGAAAGACTTGAAGTTC 1620
Oy 1621 CTTGCGGAAGCTGCTGCTAACAAGTCCGCTTCACTTCCAGTCCCTGCTGAGAGAG 1680
Db 1621 CTTGCGGAAGCTGCTGCTAACAAGTCCGCTTCACTTCCAGTCCCTGCTGAGAGAG 1680
Oy 1681 CCGCAAGGTTTGAAGCGGACCTTGTGACGCGACCGGAGCTTCACTCCGAGTACCGC 1740
Db 1681 CCGCAAGGTTTGAAGCGGACCTTGTGACGCGACCGGAGCTTCACTCCGAGTACCGC 1740
Oy 1741 GGCACCTTGACAGGACGCTTCAACCGGACGCTCAAGGCTTCAAGTACCTTGAAGAC 1800
Db 1741 GGCACCTTGACAGGACGCTTCAACCGGACGCTCAAGGCTTCAAGTACCTTGAAGAC 1800
Oy 1801 GGCCTCAAGACATGACGAGGCTCACTTCCCAAGTCAAGGCTGAGATGATGACAG 1860
Db 1801 GGCCTCAAGACATGACGAGGCTCACTTCCCAAGTCAAGGCTGAGATGATGACAG 1860
Oy 1861 GGCCTCAAGACATGACGAGGCTCACTTCCCAAGTCAAGGCTGAGATGATGACAG 1920
Db 1861 GGCCTCAAGACATGACGAGGCTCACTTCCCAAGTCAAGGCTGAGATGATGACAG 1920
Oy 1921 AAAAGACTTAACTGTGACCCCAAGAGAACTGAGAGAGGCTAATGAGACCGGCGATCAA 1980
Db 1921 AAAAGACTTAACTGTGACCCCAAGAGAACTGAGAGAGGCTAATGAGACCGGCGATCAA 1980
Oy 1981 CCGCGGTGCGCTGCGCTACCAATGTCGCTTGGATTTCTCGTGTGACACCGC 2040
Db 1981 CCGCGGTGCGCTGCGCTACCAATGTCGCTTGGATTTCTCGTGTGACACCGC 2040

```

Qy	2041	GAGGCGCCCTTTCGCTACCGTGGTTATGACGAGTCTTCACTGTTCCGCTCGTGTAC	2100
Db	2041	GAGGCGCCCTTTCGCTACCGTGGTTATGACGAGTCTTCACTGTTCCGCTCGTGTAC	2100
Qy	2101	GTCGCGATCGTGAACGCTGTTGCCGAGCTCAAGAAATAGTCTTGTAGGGAGAGTCCAC	2160
Db	2101	GTCGCGATCGTGAACGCTGTTGCCGAGCTCAAGAAATAGTCTTGTAGGGAGAGTCCAC	2160
Qy	2161	CAATCGGGTTTATAGACTTCCAAAGGACAAGCGGAAATGCGCTGCTTCCGAGGTC	2220
Db	2161	CAATCGGGTTTATAGACTTCCAAAGGACAAGCGGAAATGCGCTGCTTCCGAGGTC	2220
Qy	2221	GTTAAGCAGTGCCTGCGGCGCACTTTCAACCAAGCGCTGCTGCGCGAGTGGT	2280
Db	2221	GTTAAGCAGTGCCTGCGGCGCACTTTCAACCAAGCGCTGCTGCGCGAGTGGT	2280
Qy	2281	GCCACCGATTTTCCAGAGCTTGTACCCCGGGTGCACAACCACTCAAGGTTGGTGC	2340
Db	2281	GCCACCGATTTTCCAGAGCTTGTACCCCGGGTGCACAACCACTCAAGGTTGGTGC	2340
Qy	2341	TCCATCGACCGATCGCCCGGACCTACCGCAACAGCGGCGCAACGCTTGTGACG	2400
Db	2341	TCCATCGACCGATCGCCCGGACCTACCGCAACAGCGGCGCAACGCTTGTGACG	2400
Qy	2401	CAGGAGGAAAAGTGCAGCAACGCGGCTGAGGCGCGATGACTGTGCAAGACGAGGA	2460
Db	2401	CAGGAGGAAAAGTGCAGCAACGCGGCTGAGGCGCGATGACTGTGCAAGACGAGGA	2460
Qy	2461	CGCACTTTTGGCTGTGCTATTTGCAATTAACAAGGCTTCCACAGAGAGCAAGTCTTC	2520
Db	2461	CGCACTTTTGGCTGTGCTATTTGCAATTAACAAGGCTTCCACAGAGAGCAAGTCTTC	2520
Qy	2521	GCTGAGAGTGCACCTTCTAGTGGATCAACGCGCAACCACTGTATCATCGC	2580
Db	2521	GCTGAGAGTGCACCTTCTAGTGGATCAACGCGCAACCACTGTATCATCGC	2580
Qy	2581	GACCCGACAGGTGACATTGAGAGCACTCAACCATAGCGGAAAGCGAGTGTTC	2640
Db	2581	GACCCGACAGGTGACATTGAGAGCACTCAACCATAGCGGAAAGCGAGTGTTC	2640
Qy	2641	GACATCCCTGACACCTTGGAGATACGATCTGTAACCGAGTGAAGAGTGCAGCGAAC	2700
Db	2641	GACATCCCTGACACCTTGGAGATACGATCTGTAACCGAGTGAAGAGTGCAGCGAAC	2700
Qy	2701	GAGTGTGGCAACGATACCCCGGAGTGCACGCGGACGGAAGAAATCATCTGCTC	2760
Db	2701	GAGTGTGGCAACGATACCCCGGAGTGCACGCGGACGGAAGAAATCATCTGCTC	2760
Qy	2761	CGCAAGAACTTCCGAGGACCAACCGGATGCTGTGCTTGGCGAAGACCGGCTAC	2820
Db	2761	CGCAAGAACTTCCGAGGACCAACCGGATGCTGTGCTTGGCGAAGACCGGCTAC	2820
Qy	2821	GAGTGTGGTGGCTGCTGCTCAAAATCAAGTGAAGCTTGGCGAACCGGACCCG	2880
Db	2821	GAGTGTGGTGGCTGCTGCTCAAAATCAAGTGAAGCTTGGCGAACCGGACCCG	2880
Qy	2881	AAGCGGATAGGGGCTTCCAGAGAGGGTACAGTGGTCAAGGTCAACACGCGCTTAC	2940
Db	2881	AAGCGGATAGGGGCTTCCAGAGAGGGTACAGTGGTCAAGGTCAACACGCGCTTAC	2940
Qy	2941	AAACACGAGGCTCCAGACGCTGTTCCGCTACACCAAGCAAGCGTGAACCTCCG	3000
Db	2941	AAACACGAGGCTCCAGACGCTGTTCCGCTACACCAAGCAAGCGTGAACCTCCG	3000
Qy	3001	CTACACGAGGCTTAAAGAGAGTCAAAACGATGCTTAACTCGCTTGAACCGATTTGAG	3060
Db	3001	CTACACGAGGCTTAAAGAGAGTCAAAACGATGCTTAACTCGCTTGAACCGATTTGAG	3060
Qy	3061	TGAGCTGTCACTGAAGACGCGGCTGACGAGCTGTCTTTCGAGACCGAGCTCAATTCAC	3120
Db	3061	TGAGCTGTCACTGAAGACGCGGCTGACGAGCTGTCTTTCGAGACCGAGCTCAATTCAC	3120
Qy	3121	CAACGCGGCGGACCGTGTGAAGACCTGTGAGGACGACGACCCCTCATCTCGTACATA	3180
Db	3121	CAACGCGGCGGACCGTGTGAAGACCTGTGAGGACGACGACCCCTCATCTCGTACATA	3180
Qy	3181	GACTCCCTTATGAAGACTCAGCAAGAAAGTGTCCGCCAAGCGATCAATAGGGCAAGTTC	3240
Db	3181	GACTCCCTTATGAAGACTCAGCAAGAAAGTGTCCGCCAAGCGATCAATAGGGCAAGTTC	3240
Qy	3241	GAGCAGGGAGTCCCGCTCACTCAAAAGTCTCTCAACTTCCGCTGCGGCTTGGATACG	3300
Db	3241	GAGCAGGGAGTCCCGCTCACTCAAAAGTCTCTCAACTTCCGCTGCGGCTTGGATACG	3300
Qy	3301	ATACTGAGAGATCTCCGATCCGAGGCGGACGCGATCGGTACAGCAACGCTTCCC	3360
Db	3301	ATACTGAGAGATCTCCGATCCGAGGCGGACGCGATCGGTACAGCAACGCTTCCC	3360
Qy	3361	GACGAGGAGAGGCGCATGCTGCTGAGAGGGAATCAATCAATGCCACACGCGATTC	3420
Db	3361	GACGAGGAGAGGCGCATGCTGCTGAGAGGGAATCAATCAATGCCACACGCGATTC	3420
Qy	3421	GTCGCGGAGCTGAGACCGAGTTGACACCGCCCAATAACAGAGTGAAGTCTTTC	3480
Db	3421	GTCGCGGAGCTGAGACCGAGTTGACACCGCCCAATAACAGAGTGAAGTCTTTC	3480
Qy	3481	GCCGCTTTTAAAGCGCATCGGACGCTGACAGCTGCTGTAATCTAATCAGAGAACG	3540
Db	3481	GCCGCTTTTAAAGCGCATCGGACGCTGACAGCTGCTGTAATCTAATCAGAGAACG	3540
Qy	3541	TGTGGGAAAGCGACCTTGGAGGAGGAGGCTTGGCTCCGTTGAAGTGCAGGCTGCTC	3600
Db	3541	TGTGGGAAAGCGACCTTGGAGGAGGAGGCTTGGCTCCGTTGAAGTGCAGGCTGCTC	3600
Qy	3601	GACTCCGCGACGCTTGAACGCTTGGCGGCAACCACTTCTCTGCGCGGCTGATGCTC	3660
Db	3601	GACTCCGCGACGCTTGAACGCTTGGCGGCAACCACTTCTCTGCGCGGCTGATGCTC	3660
Qy	3661	AGCCTCTTCCGCGGCGTCAAGTTTCGAGCTTTCGAGGCGGACGCTGCTGCTGAGT	3720
Db	3661	AGCCTCTTCCGCGGCGTCAAGTTTCGAGCTTTCGAGGCGGACGCTGCTGCTGAGT	3720
Qy	3721	AGCCTTAACTCCGTTTCCAGCGCTGACGCGGCTTCAATGAGGCGAAGCTTACAGACCAA	3780
Db	3721	AGCCTTAACTCCGTTTCCAGCGCTGACGCGGCTTCAATGAGGCGAAGCTTACAGACCAA	3780
Qy	3781	CATTGAAGGTGAGGGTGCAGAAATGTCGCGGACATCGGACCTCCGCTGCTGAG	3840
Db	3781	CATTGAAGGTGAGGGTGCAGAAATGTCGCGGACATCGGACCTCCGCTGCTGAG	3840
Qy	3841	CAGTGTCTCTGACACCTGTCAAGAGGCTCTCAAGATATTTGGGCGCTGTACAAAGC	3900
Db	3841	CAGTGTCTCTGACACCTGTCAAGAGGCTCTCAAGATATTTGGGCGCTGTACAAAGC	3900
Qy	3901	GAACTCTTAACTCAAGTACGTGAGAGGCTGTGAGAGATCACCAAGGCTGAGTGC	3960
Db	3901	GAACTCTTAACTCAAGTACGTGAGAGGCTGTGAGAGATCACCAAGGCTGAGTGC	3960
Qy	3961	GCCGCTACCAAGGCTCTGTGCAATGTACAGATGCAATCAATTAACGCGCGGAG	4020
Db	3961	GCCGCTACCAAGGCTCTGTGCAATGTACAGATGCAATCAATTAACGCGCGGAG	4020
Qy	4021	TTTGGCGGCTATCATGACGCTGTTGTTCCGCTTGGGCGGCGGACTTCCGTTTGA	4080
Db	4021	TTTGGCGGCTATCATGACGCTGTTGTTCCGCTTGGGCGGCGGACTTCCGTTTGA	4080
Qy	4081	CAACTGCGCTGTGTGCTGCTCAATGTGCAAGGACCGGACGCTTACAGACACGATTCG	4140
Db	4081	CAACTGCGCTGTGTGCTGCTCAATGTGCAAGGACCGGACGCTTACAGACACGATTCG	4140
Qy	4141	GCTAACGTCGCGCATGCTGCTTGAACAGTCTTTCGAGGCGGCGGCGGCGGCGG	4200
Db	4141	GCTAACGTCGCGCATGCTGCTTGAACAGTCTTTCGAGGCGGCGGCGGCGGCGGCGG	4200
Qy	4201	GCAAGTTTCTGTTGCAATGTGCGAAGCGGAAACGCTTCTTCACTTACCGGAAAGCT	4260
Db	4201	GCAAGTTTCTGTTGCAATGTGCGAAGCGGAAACGCTTCTTCACTTACCGGAAAGCT	4260

Db 4201 GCAAGTTTCGTTGCGACATGTCGGAAGCCGAAAGCCCTTCTTCACTACCGCGAAAGCT 4260
 Oy 4261 GGTGTTTCTGCGACATCAAGCCACGTTGCGACTGAGACTGCGCCCGGAGTCTTCATG 4320
 Db 4261 GGTGTTTCTGCGACATCAAGCCACGTTGCGACTGAGACTGCGCCCGGAGTCTTCATG 4320
 Oy 4321 GATGCACTGCGACCAAGCTTTTCGAGATTAATGACACCGGANAACCCCGTCCATCA 4380
 Db 4321 GATGCACTGCGACCAAGCTTTTCGAGATTAATGACACCGGANAACCCCGTCCATCA 4380
 Oy 4381 TCCTGCGCGTCACTGCTTTCATCGACTCTCTACATGATGAGAGGTGCTGAGT 4440
 Db 4381 TCCTGCGCGTCACTGCTTTCATCGACTCTCTACATGATGAGAGGTGCTGAGT 4440
 Oy 4441 GGAGACACCGGACGACAGAGACCTTGAACAGACGAAAGCCGCTTTCGACAGACAG 4500
 Db 4441 GGAGACACCGGACGACAGAGACCTTGAACAGACGAAAGCCGCTTTCGACAGACAG 4500
 Oy 4501 CAATGACGCTGCTGGAATGCTGACAGAGAGGAGAAAGACAGGACGATTCCTTAAT 4560
 Db 4501 CAATGACGCTGCTGGAATGCTGACAGAGAGGAGAAAGACAGGACGATTCCTTAAT 4560
 Oy 4561 GCGCCGACGCTGCTGAGAGCCGCTCATTTCTCATTTTCGAAAGAGTCTGAGTGGACG 4620
 Db 4561 GCGCCGACGCTGCTGAGAGCCGCTCATTTCTCATTTTCGAAAGAGTCTGAGTGGACG 4620
 Oy 4621 CGGACTGTGCGCGCTGCGACATCACCTTGCGAACCCCATCTGCGGACGAGGTT 4680
 Db 4621 CGGACTGTGCGCGCTGCGACATCACCTTGCGAACCCCATCTGCGGACGAGGTT 4680
 Oy 4681 GCGCGGAGACCAAGCCCGGACTTTTGCGCTTCTCTGAGTCCCATGTCGCGCTGTC 4740
 Db 4681 GCGCGGAGACCAAGCCCGGACTTTTGCGCTTCTCTGAGTCCCATGTCGCGCTGTC 4740
 Oy 4741 TCGAAGCGGTAACGCGCCCGGACGACTGCGCTTTCGAAAGAGTCAACCCGCTCCACG 4800
 Db 4741 TCGAAGCGGTAACGCGCCCGGACGACTGCGCTTTCGAAAGAGTCAACCCGCTCCACG 4800
 Oy 4801 TGGAGGCGGTGACCGGACCGGACCGGAGTCAAGGAGGAGCCGAGACAGCGCGGTC 4860
 Db 4801 TGGAGGCGGTGACCGGACCGGACCGGAGTCAAGGAGGAGCCGAGACAGCGCGGTC 4860
 Oy 4861 GTCCAGGCTCTGATCAGCGGCGCTTATCTCTCAAGAGACGAGCTTCTCCGACGATCC 4920
 Db 4861 GTCCAGGCTCTGATCAGCGGCGCTTATCTCTCAAGAGACGAGCTTCTCCGACGATCC 4920
 Oy 4921 AAAAGGTACTCAAGACTAAAGGAGTGTCAATCCACTCTTTCTCTGCGCGAGTGG 4980
 Db 4921 AAAAGGTACTCAAGACTAAAGGAGTGTCAATCCACTCTTTCTCTGCGCGAGTGG 4980
 Oy 4981 GATTACAGGCGCGGACTGCGACAGTCCGAGTCTGCGCGCGCTGCGAGAGTGG 5040
 Db 4981 GATTACAGGCGCGGACTGCGACAGTCCGAGTCTGCGCGCGCTGCGAGAGTGG 5040
 Oy 5041 CGCTCATGTAATTCAGAGCGCTTGTGCTTATCTGCGCGACAGTGGCCACTTGAAGCGCAT 5100
 Db 5041 CGCTCATGTAATTCAGAGCGCTTGTGCTTATCTGCGCGACAGTGGCCACTTGAAGCGCAT 5100
 Oy 5101 CGCTCATGTAATTCAGAGCGCTTGTGCTTATCTGCGCGACAGTGGCCACTTGAAGCGCAT 5160
 Db 5101 CGCTCATGTAATTCAGAGCGCTTGTGCTTATCTGCGCGACAGTGGCCACTTGAAGCGCAT 5160
 Oy 5161 ATTAGTTTCTGTTGTAACAGAGTGGTCCCTCCCATTTGAGTAAAGACTCTGTGAG 5220
 Db 5161 ATTAGTTTCTGTTGTAACAGAGTGGTCCCTCCCATTTGAGTAAAGACTCTGTGAG 5220
 Oy 5221 TCTCTCAAGCTTACTGTTGAGTCTGCTGCGGTTCAATTCATTCGCAAGCAAGAGGT 5280
 Db 5221 TCTCTCAAGCTTACTGTTGAGTCTGCTGCGGTTCAATTCATTCGCAAGCAAGAGGT 5280
 Oy 5281 GCGCAACTAGTACGCGCGCCCTGGAGTACA 5312
 Db 5281 GCGCAACTAGTACGCGCGCCCTGGAGTACA 5312

RESULT 2
 AR125093
 LOCUS AR125093 5312 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 41 from patent US 6,177,075.
 AR125093
 ACCESSION AR125093
 VERSION AR125093.1 GI:14111155
 KEYWORDS
 SOURCE
 ORGANISM
 UNKNOWN.
 REFERENCE
 1 (bases 1 to 5312)
 AUTHORS
 Christlian,P.Daniel., Gordon,K.Henrich,Julius, and
 Hanzlik,T.Nelson.
 TITLE
 Insect viruses and their uses in protecting plants
 JOURNAL
 Patent: US 6,177,075-A 41 23-JAN-2001;
 FEATURES
 source
 1..5312
 /organism="unknown"
 BASE COUNT 1133 a 1740 c 1432 g 1007 t
 ORIGIN
 Query Match 100.0%; Score 5312; DB 6; Length 5312;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1 GTTCTGCTCCCGCGAGCGTAAATATAGGGGAACAATGTACGCGAAAGCAGACGCTG 60
 Db 1 GTTCTGCTCCCGCGAGCGTAAATATAGGGGAACAATGTACGCGAAAGCAGACGCTG 60
 Oy 61 GCGGTGTCTACGCGCGGAGATGTGCTTACGCGAAGTATCTGACAGAGAGCAATC 120
 Db 61 GCGGTGTCTACGCGCGGAGATGTGCTTACGCGAAGTATCTGACAGAGAGCAATC 120
 Oy 61 GCGGTGTCTACGCGCGGAGATGTGCTTACGCGAAGTATCTGACAGAGAGCAATC 120
 Db 61 GCGGTGTCTACGCGCGGAGATGTGCTTACGCGAAGTATCTGACAGAGAGCAATC 120
 Oy 121 AAGTTGACTTGCAGCGCGGACGTAAGGACCTGACAGACTGACTATCTCG 180
 Db 121 AAGTTGACTTGCAGCGCGGACGTAAGGACCTGACAGACTGACTATCTCG 180
 Oy 121 AAGTTGACTTGCAGCGCGGACGTAAGGACCTGACAGACTGACTATCTCG 180
 Db 121 AAGTTGACTTGCAGCGCGGACGTAAGGACCTGACAGACTGACTATCTCG 180
 Oy 181 CTGCGCTTCAAGGGGGGACTTTTACCCCGGACACAAACCCGATCTGCGCGGACCAA 240
 Db 181 CTGCGCTTCAAGGGGGGACTTTTACCCCGGACACAAACCCGATCTGCGCGGACCAA 240
 Oy 181 CTGCGCTTCAAGGGGGGACTTTTACCCCGGACACAAACCCGATCTGCGCGGACCAA 240
 Db 181 CTGCGCTTCAAGGGGGGACTTTTACCCCGGACACAAACCCGATCTGCGCGGACCAA 240
 Oy 241 CGTGTGCGAAGAGGTTCTGCACAATTTCCGAGGGGAGTGAAGACAGTGTCTGAGATA 300
 Db 241 CGTGTGCGAAGAGGTTCTGCACAATTTCCGAGGGGAGTGAAGACAGTGTCTGAGATA 300
 Oy 241 CGTGTGCGAAGAGGTTCTGCACAATTTCCGAGGGGAGTGAAGACAGTGTCTGAGATA 300
 Db 241 CGTGTGCGAAGAGGTTCTGCACAATTTCCGAGGGGAGTGAAGACAGTGTCTGAGATA 300
 Oy 301 GGGCGGCTCTGCAAGGCGACTTAAAGTACATGGGGGACCGAAAGCCCGGTCGACAGC 360
 Db 301 GGGCGGCTCTGCAAGGCGACTTAAAGTACATGGGGGACCGAAAGCCCGGTCGACAGC 360
 Oy 301 GGGCGGCTCTGCAAGGCGACTTAAAGTACATGGGGGACCGAAAGCCCGGTCGACAGC 360
 Db 301 GGGCGGCTCTGCAAGGCGACTTAAAGTACATGGGGGACCGAAAGCCCGGTCGACAGC 360
 Oy 361 TATCAAGGAGTGCACCAAGTACAGGACCCGCGAGCTGCGACATTAAGGCTTAAAG 420
 Db 361 TATCAAGGAGTGCACCAAGTACAGGACCCGCGAGCTGCGACATTAAGGCTTAAAG 420
 Oy 361 TATCAAGGAGTGCACCAAGTACAGGACCCGCGAGCTGCGACATTAAGGCTTAAAG 420
 Db 361 TATCAAGGAGTGCACCAAGTACAGGACCCGCGAGCTGCGACATTAAGGCTTAAAG 420
 Oy 421 TCTAATCTGCTGCGACAGCGCGGCGGCGAGTCAAGGCGGACGCTCACTGCTCCCAAC 480
 Db 421 TCTAATCTGCTGCGACAGCGCGGCGGCGAGTCAAGGCGGACGCTCACTGCTCCCAAC 480
 Oy 421 TCTAATCTGCTGCGACAGCGCGGCGGCGAGTCAAGGCGGACGCTCACTGCTCCCAAC 480
 Db 421 TCTAATCTGCTGCGACAGCGCGGCGGCGAGTCAAGGCGGACGCTCACTGCTCCCAAC 480
 Oy 481 GGCATTGCTCCGCGACCTTGTGCTGCGAGGAGTGGCTCTTGGCGTTCAATTCGCGC 540
 Db 481 GGCATTGCTCCGCGACCTTGTGCTGCGAGGAGTGGCTCTTGGCGTTCAATTCGCGC 540
 Oy 481 GGCATTGCTCCGCGACCTTGTGCTGCGAGGAGTGGCTCTTGGCGTTCAATTCGCGC 540
 Db 481 GGCATTGCTCCGCGACCTTGTGCTGCGAGGAGTGGCTCTTGGCGTTCAATTCGCGC 540
 Oy 541 GTTGAATTTGCCAATCACTCCCTCTATGACGTGACCTTAGAGAGCTGGCCAAATCGTTT 600
 Db 541 GTTGAATTTGCCAATCACTCCCTCTATGACGTGACCTTAGAGAGCTGGCCAAATCGTTT 600
 Oy 541 GTTGAATTTGCCAATCACTCCCTCTATGACGTGACCTTAGAGAGCTGGCCAAATCGTTT 600
 Db 541 GTTGAATTTGCCAATCACTCCCTCTATGACGTGACCTTAGAGAGCTGGCCAAATCGTTT 600
 Oy 601 GAGAACACAGGACTTCAATGTGCTGCGGCTTATGACATGCGCAAGAGAGTGTCTTAC 660
 Db 601 GAGAACACAGGACTTCAATGTGCTGCGGCTTATGACATGCGCAAGAGAGTGTCTTAC 660
 Oy 601 GAGAACACAGGACTTCAATGTGCTGCGGCTTATGACATGCGCAAGAGAGTGTCTTAC 660
 Db 601 GAGAACACAGGACTTCAATGTGCTGCGGCTTATGACATGCGCAAGAGAGTGTCTTAC 660
 Oy 661 ATGGAACAGGAGTGAATGCGAGCTGAGCTGCGCTTCAAGTATTGAAGAGCTATG 720
 Db 661 ATGGAACAGGAGTGAATGCGAGCTGAGCTGCGCTTCAAGTATTGAAGAGCTATG 720

Dh 661 ATGACAAAGCTGGTAAATGCGAGCTCGGCTACCGCTTCCAGTTATTGAAGAGCTATG 720
Qy 721 GCTGTGAAGACTGCGGATTCAGAGGGGGGAGACTCCGTCTCCACTTCCCTGAATTGAG 780
Dh 722 GCTGTGAAGACTGCGGATTCAGAGGGGGGAGACTCCGTCTCCACTTCCCTGAATTGAG 780
Qy 781 TTATCAAGAGAGCCAAAGAGCGGCGATCGAGAGGCTGGCGCGCGCGGCTCTACTCC 840
Dh 781 TTATCAAGAGAGCCAAAGAGCGGCGATCGAGAGGCTGGCGCGCGCGGCTCTACTCC 840
Qy 841 AGAGCGCGCTCAATTTCTCGCGGAGAGAGAGCTGGGGTGAATGGTACTTACAGACTTC 900
Dh 841 AGAGCGCGCTCAATTTCTCGCGGAGAGAGAGCTGGGGTGAATGGTACTTACAGACTTC 900
Qy 901 CACACATGAGCTCGCTACTACTGTGTAGAGAACTAACCCCACTCCGTTGGTTTCTCATC 960
Dh 901 CACACATGAGCTCGCTACTACTGTGTAGAGAACTAACCCCACTCCGTTGGTTTCTCATC 960
Qy 961 CATATTAAGTCCAGAGCGCGCAAGGCTCCAGCAATTAGCTGGCGATCACTCGCGCGCA 1020
Dh 961 CATATTAAGTCCAGAGCGCGCAAGGCTCCAGCAATTAGCTGGCGATCACTCGCGCGCA 1020
Qy 1021 CTTGAGAGCGGATGCTGCGCGCTGCTCCAAAGAGCTCCCAAGGCTCTGCGAGATCCCA 1080
Dh 1021 CTTGAGAGCGGATGCTGCGCGCTGCTCCAAAGAGCTCCCAAGGCTCTGCGAGATCCCA 1080
Qy 1081 AACATCTTTTATTAAGCGCGAGCGCTGCGGCACTGAGCATTAAGACATCTTACCTCAAG 1140
Dh 1081 AACATCTTTTATTAAGCGCGAGCGCTGCGGCACTGAGCATTAAGACATCTTACCTCAAG 1140
Qy 1141 CACAAATCAACATGCTGCTCAATTTATGCAAAAGGCTCCTGGAAGAGAACTAGTGAC 1200
Dh 1141 CACAAATCAACATGCTGCTCAATTTATGCAAAAGGCTCCTGGAAGAGAACTAGTGAC 1200
Qy 1201 ATGACCGCTTGAATGTGCTGCGCGCTAGAGCTGGCGCGAGATCGTGTGCTCTGAA 1260
Dh 1201 ATGACCGCTTGAATGTGCTGCGCGCTAGAGCTGGCGCGAGATCGTGTGCTCTGAA 1260
Qy 1261 GTACCGGAGAGCTCTGGAACATCTCACCGGCTGACCTGCTGCTGCTGCTGCT 1320
Dh 1261 GTACCGGAGAGCTCTGGAACATCTCACCGGCTGACCTGCTGCTGCTGCTGCT 1320
Qy 1321 TACGCTCTCAACATCACTGAGCGCGGAGAGGCTGGCTGCTGCTGCTGCTGCTGCT 1380
Dh 1321 TACGCTCTCAACATCACTGAGCGCGGAGAGGCTGGCTGCTGCTGCTGCTGCTGCT 1380
Qy 1381 GAGCTCTTTGAGAGAGCTTGGTCTGGAGAGTCTCAAGCAGCTTGGGCTCTGTTGC 1440
Dh 1381 GAGCTCTTTGAGAGAGCTTGGTCTGGAGAGTCTCAAGCAGCTTGGGCTCTGTTGC 1440
Qy 1441 GGTCTGCGCAACTCAAAAGGAGCGAGCTGCTTTACTTAAGCGGCTGCTGATTAAGTAC 1500
Dh 1441 GGTCTGCGCAACTCAAAAGGAGCGAGCTGCTTTACTTAAGCGGCTGCTGATTAAGTAC 1500
Qy 1501 CGAGTCACTGCTGCTGCGAGAGATTAATCTGCGAGCTCGCGCTGCTGCTGCTGCTGCT 1560
Dh 1501 CGAGTCACTGCTGCTGCGAGAGATTAATCTGCGAGCTCGCGCTGCTGCTGCTGCTGCT 1560
Qy 1561 TTCTGCGCGTCCGCGTACCACTGCGCGCTCTTTCAAGAGAGAGAGCTTGAAGTTC 1620
Dh 1561 TTCTGCGCGTCCGCGTACCACTGCGCGCTCTTTCAAGAGAGAGAGCTTGAAGTTC 1620
Qy 1621 CTTGCGAGAGCTGCTGCTCAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Dh 1621 CTTGCGAGAGCTGCTGCTCAAGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
Qy 1681 CCGCAAGATTTGAGAGCGCGAGATTTGAGAGCGCGAGCTGCTGCTGCTGCTGCTGCT 1740
Dh 1681 CCGCAAGATTTGAGAGCGCGAGATTTGAGAGCGCGAGCTGCTGCTGCTGCTGCTGCT 1740
Qy 1741 GCGACCTTGGAGAGAGTCTCAACAGAGCTCAAGAGAGCTCAAGATCACTCTGAGAGAC 1800
Dh 1741 GCGACCTTGGAGAGAGTCTCAACAGAGCTCAAGAGAGCTCAAGATCACTCTGAGAGAC 1800

Qy 1801 GCCCTCAAGACATCGAGCGGAGCTCACTCTCCCACTCAAGAGCGCTCGAGATGTACAG 1860
Dh 1801 GCCCTCAAGACATCGAGCGGAGCTCACTCTCCCACTCAAGAGCGCTCGAGATGTACAG 1860
Qy 1861 GGCCTGCGAGAGCGGAG 1920
Dh 1861 GGCCTGCGAGAGCGGAG 1920
Qy 1921 AAAGCATTTACGTGAG 1980
Dh 1921 AAAGCATTTACGTGAG 1980
Qy 1981 CCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
Dh 1981 CCGCGCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2040
Qy 2041 GAGGAGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
Dh 2041 GAGGAGCGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100
Qy 2101 GTCGAGTCTGAG 2160
Dh 2101 GTCGAGTCTGAG 2160
Qy 2161 CAAATCGGCTTTATGACTTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Dh 2161 CAAATCGGCTTTATGACTTCCAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220
Qy 2221 GTTAAAG 2280
Dh 2221 GTTAAAG 2280
Qy 2281 GCCACAGCTTTTCCAG 2340
Dh 2281 GCCACAGAGTCTTCCAG 2340
Qy 2341 TCCATCAGCAG 2400
Dh 2341 TCCATCAGCAG 2400
Qy 2401 CAG 2460
Dh 2401 CAG 2460
Qy 2461 CGACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
Dh 2461 CGACCTTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2520
Qy 2521 GCTGAG 2580
Dh 2521 GCTGAG 2580
Qy 2581 GACCCGAG 2640
Dh 2581 GACCCGAG 2640
Qy 2641 GACATCCCTGAG 2700
Dh 2641 GACATCCCTGAG 2700
Qy 2701 GAAATGAG 2760
Dh 2701 GAAATGAG 2760
Qy 2761 CGCAAG 2820
Dh 2761 CGCAAG 2820
Qy 2821 GAGGTGTTGGCGGTGCTGCAAAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880
Dh 2821 GAGGTGTTGGCGGTGCTGCAAAATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2880

QY 2881 AAGCGCATAGGCGCTTCAGGAAGGGGTACAGTGGGTCAAGGTCAACCAACGCGCTAC 2940
 Db 2881 AAGCGCATAGGCGCTTCAGGAAGGGGTACAGTGGGTCAAGGTCAACCAACGCGCTAC 2940
 QY 2941 AAACACAGAGCGCTCAGACGCTGTGTCCGCTACACCAAGCGAAGCGTACCTGCG 3000
 Db 2941 AAACACAGAGCGCTCAGACGCTGTGTCCGCTACACCAAGCGAAGCGTACCTGCG 3000
 QY 3001 CTACAGGAAGCTAAGAGAGAGCTCAACGCGATGTCTAACTCGCTTGAACCGACATTTGGAC 3060
 Db 3001 CTACAGGAAGCTAAGAGAGAGCTCAACGCGATGTCTAACTCGCTTGAACCGACATTTGGAC 3060
 QY 3061 TGGACTGTCACTGAAGAGCGCCGCTGACCGAGCTGCTTGAAGACCGAGTCAAGTTCAAC 3120
 Db 3061 TGGACTGTCACTGAAGAGCGCCGCTGACCGAGCTGCTTGAAGACCGAGTCAAGTTCAAC 3120
 QY 3121 CAACGCGCGGACCGCTGCAAGAGCTGTGAGAGCGACGACCCCTTACCTCGTACATA 3180
 Db 3121 CAACGCGCGGACCGCTGCAAGAGCTGTGAGAGCGACGACCCCTTACCTCGTACATA 3180
 QY 3181 GACTTCTTATGAGAGCTTCAGAGAAAGTGTGCGCCCAAGCGATCAATACGGGCAAGGTC 3240
 Db 3181 GACTTCTTATGAGAGCTTCAGAGAAAGTGTGCGCCCAAGCGATCAATACGGGCAAGGTC 3240
 QY 3241 GGGGAGGGGATGCGCGCTTCACTCAAGTCTCAACTTGTGCTGCGCGCTTGAATAGG 3300
 Db 3241 GGGGAGGGGATGCGCGCTTCACTCAAGTCTCAACTTGTGCTGCGCGCTTGAATAGG 3300
 QY 3301 ATACTGAGAGAGATCTCCGTACCGGAGCGCGACGCTCGGTAGACGAACGGTCTCCG 3360
 Db 3301 ATACTGAGAGAGATCTCCGTACCGGAGCGCGACGCTCGGTAGACGAACGGTCTCCG 3360
 QY 3361 GACGAGAAAGAGGCAATGTCTGCTCAAGCGAAAGATCAATCAATGCCACACCGCACGTT 3420
 Db 3361 GACGAGAAAGAGGCAATGTCTGCTCAAGCGAAAGATCAATCAATGCCACACCGCACGTT 3420
 QY 3421 GTCTCGGCGGATCTGACCCGAGTTTGAACCGCGCAATCAATCAAGTGAAGTGGCTCTTC 3480
 Db 3421 GTCTCGGCGGATCTGACCCGAGTTTGAACCGCGCAATCAATCAAGTGAAGTGGCTCTTC 3480
 QY 3481 GCGCGCTTTTGAAGCGCATCGGACAGCTGACAGCTGCGCTTAACTTATTCAGAGAACG 3540
 Db 3481 GCGCGCTTTTGAAGCGCATCGGACAGCTGACAGCTGCGCTTAACTTATTCAGAGAACG 3540
 QY 3541 TGTGGAAGAGCGACCTTGGAGGGAAGGTCTAGGCTCGTGAAGTGAAGTGTGCTGTC 3600
 Db 3541 TGTGGAAGAGCGACCTTGGAGGGAAGGTCTAGGCTCGTGAAGTGAAGTGTGCTGTC 3600
 QY 3601 GACTCGGCGGAGCTTGAAGCGCTTGGCGCAACCATCTTCTGCGCGCGCTCATGTC 3660
 Db 3601 GACTCGGCGGAGCTTGAAGCGCTTGGCGCAACCATCTTCTGCGCGCGCTCATGTC 3660
 QY 3661 AGCGCTTTCGAGCGGCTCAAGTTGCGAGCTTCAAGAGCGAGCATCGCTCTGCTG 3720
 Db 3661 AGCGCTTTCGAGCGGCTCAAGTTGCGAGCTTCAAGAGCGAGCATCGCTCTGCTG 3720
 QY 3721 AGCGATTAACCTCGTTTGAAGCGTACCGCTTCAATGAGCGAAGCTTCAAGAACAA 3780
 Db 3721 AGCGATTAACCTCGTTTGAAGCGTACCGCTTCAATGAGCGAAGCTTCAAGAACAA 3780
 QY 3781 CATTTGAAGTGAAGTGAAGAAATGCGCGTATCATGGAATCTCTGCTGCTGAG 3840
 Db 3781 CATTTGAAGTGAAGTGAAGAAATGCGCGTATCATGGAATCTCTGCTGCTGAG 3840
 QY 3841 CAGGTGTCTGAGACCTGTGAGAGCGCTCAAGATATTTGGGCGCTGCTACAAAG 3900
 Db 3841 CAGGTGTCTGAGACCTGTGAGAGCGCTCAAGATATTTGGGCGCTGCTACAAAG 3900
 QY 3901 GAACTCTTTACTCAAGTACGTGAGGCTGTGAGAGACATCAACAAAGGCTGAGTGAAC 3960
 Db 3901 GAACTCTTTACTCAAGTACGTGAGGCTGTGAGAGACATCAACAAAGGCTGAGTGAAC 3960
 QY 3961 GCCCGTACCAAGCTCTGTGCGCACATGTCAAGCATGTACTCAATTAACGCGCGAG 4020

Db 3961 GCCCGTACCAAGCTCTGTGCGCACATGTCAAGCATGTACTCAATTAACGCGCGAG 4020
 QY 4021 TGTGCGGATGATCATCATGACGCTGTTGTGCTTTGGGCGCGGACCTTCCGTTGAA 4080
 Db 4021 TGTGCGGATGATCATCATGACGCTGTTGTGCTTTGGGCGCGGACCTTCCGTTGAA 4080
 QY 4081 CAACGCGCGGATGATCATCATGACGCTGTTGTGCTTTGGGCGCGGACCTTCCGTTGAA 4140
 Db 4081 CAACGCGCGGATGATCATCATGACGCTGTTGTGCTTTGGGCGCGGACCTTCCGTTGAA 4140
 QY 4141 GCTAACGTCGCGGATGATCATCATGACGCTTTCAGAGCGCCGCGACGCGCGCG 4200
 Db 4141 GCTAACGTCGCGGATGATCATCATGACGCTTTCAGAGCGCCGCGACGCGCGCG 4200
 QY 4201 GCGAGTTCTGTTGCGACATGTGCGAAGCGGAAAGCGCTTTCATCTTACCGGAAAGCT 4260
 Db 4201 GCGAGTTCTGTTGCGACATGTGCGAAGCGGAAAGCGCTTTCATCTTACCGGAAAGCT 4260
 QY 4261 GGTGTTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
 Db 4261 GGTGTTCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4320
 QY 4321 GATGACCTGACGCGCAACAGCTTTCGAGATTTTGAACCGGAGACCGCGTCAATCA 4380
 Db 4321 GATGACCTGACGCGCAACAGCTTTCGAGATTTTGAACCGGAGACCGCGTCAATCA 4380
 QY 4381 TCTTCGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4440
 Db 4381 TCTTCGCGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4440
 QY 4441 GAGAGACCGGAGACCAAGAGACCTTGAACGAGAGAGCGCGCTTCCGAGACAGG 4500
 Db 4441 GAGAGACCGGAGACCAAGAGACCTTGAACGAGAGAGCGCGCTTCCGAGACAGG 4500
 QY 4501 CAATACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560
 Db 4501 CAATACGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4560
 QY 4561 GCGCCGCTGCTCGAGCGCGCTTCAATCTTTCGAAAGAGCTGACGCGGACCGG 4620
 Db 4561 GCGCCGCTGCTCGAGCGCGCTTCAATCTTTCGAAAGAGCTGACGCGGACCGG 4620
 QY 4621 CGGATGTCGCGCTGCGACATCACTTCCGCAACCCCATCTGCGGACCGGACAGGTT 4680
 Db 4621 CGGATGTCGCGCTGCGACATCACTTCCGCAACCCCATCTGCGGACCGGACAGGTT 4680
 QY 4681 GCGCGGAGACCAAGCGGAGCTTTCGCGCTTTCGAGGTTCCAGTCTGCGGCTGTC 4740
 Db 4681 GCGCGGAGACCAAGCGGAGCTTTCGCGCTTTCGAGGTTCCAGTCTGCGGCTGTC 4740
 QY 4741 TCGAAGCGGTACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4800
 Db 4741 TCGAAGCGGTACCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4800
 QY 4801 TGGAAAGGCTGACCGGAGACCGGAGAGTCAAGGAGGAGACCGGAGACAGCGCGGTC 4860
 Db 4801 TGGAAAGGCTGACCGGAGACCGGAGAGTCAAGGAGGAGACCGGAGACAGCGCGGTC 4860
 QY 4861 GTGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4920
 Db 4861 GTGCGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 4920
 QY 4921 AAAGGCTACTCAAGAACTAAGGATGCTCAACATCACTTTCGAGCGCGAGTGC 4980
 Db 4921 AAAGGCTACTCAAGAACTAAGGATGCTCAACATCACTTTCGAGCGCGAGTGC 4980
 QY 4981 GATTACAGGCGCGGAGCTGCGGACAGTTCGAGGCTGCGGCGCGGCGGAGATGAGG 5040
 Db 4981 GATTACAGGCGCGGAGCTGCGGACAGTTCGAGGCTGCGGCGCGGCGGAGATGAGG 5040
 QY 5041 CGCTCATGATTCAGAGCGCGGCTTCAATGCGCGGACAGTGCAGCTTGAAGGACATA 5100

Db 5041 CGCTCATGTATTCACGAGCGGTTGGCTTCACTCTGCCGCGAGTCCGACCTTGAAGCGGATA 5100
 Qy 5101 CGCTTACCTCGAGCTCTGTTCCCGATGTAAAGATCAGCAAGAGCGCATGAAGAACAA 5160
 Db 5101 CGCTTACCTCGAGCTCTGTTCCCGATGTAAAGATCAGCAAGAGCGCATGAAGAACAA 5160
 Qy 5161 ATTAGTTCTCTGTTGGTAAACAGAGTGTGTCCCTCCATTGAGGTAAAGACTCTGGTGGAG 5220
 Db 5161 ATTAGTTCTCTGTTGGTAAACAGAGTGTGTCCCTCCATTGAGGTAAAGACTCTGGTGGAG 5220
 Qy 5221 TCCTCAACGTTACTGTTAGTGTCTGCGGTTGATTCATTCCTCAAGCAGCAAAAGGT 5280
 Db 5221 TCCTCAACGTTACTGTTAGTGTCTGCGGTTGATTCATTCCTCAAGCAGCAAAAGGT 5280
 Qy 5281 GCGCACTAGTACGCGCGCCCTTGGGATACCA 5312
 Db 5281 GCGCACTAGTACGCGCGCCCTTGGGATACCA 5312
 RESULT 3
 ARI25094
 LOCUS ARI25094 5312 bp DNA linear PAT 16-MAY-2001
 DEFINITION Sequence 43 from patent US 6177075.
 ACCESSION ARI25094
 VERSION ARI25094.1 GI:14111156
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 5312)
 AUTHORS Christian P. Daniel., Gordon, K. Hienrich, Julius, and
 Hamzlik, T. Nelson.
 TITLE Insect viruses and their uses in protecting plants
 JOURNAL Patent: US 6177075-A 43 23-JAN-2001;
 FEATURES
 source 1..5312
 location/Qualifiers
 BASE COUNT 1133 a 1740 c 1432 g 1007 t
 ORIGIN
 Query Match 100.0%; Score 5312; DB 6; Length 5312;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 421 TCTAGATCCGTCGCCACAGGCCGCGCCCGAGTTCAAGGCGGACGCTTCACTGCTGCCAAC 480
 Qy 481 GGCAATTCCTCCGCGACCTTTCTGCGTGAAGAGTGGGCTTTGGCGGTTCAATCGCGC 540
 Db 481 GGCAATTCCTCCGCGACCTTTCTGCGTGAAGAGTGGGCTTTGGCGGTTCAATCGCGC 540
 Qy 541 GTTGAATTTGCCAATCACTCCCTCTATGAGAGTAAACCTTAAGAGAGTGGGCAATGGGTTT 600
 Db 541 GTTGAATTTGCCAATCACTCCCTCTATGAGAGTAAACCTTAAGAGAGTGGGCAATGGGTTT 600
 Qy 601 GAGAACCGAGGACTTCAATGTCGCGCGGCTTATGACACATGCGAGAAAGTCTCTAC 660
 Db 601 GAGAACCGAGGACTTCAATGTCGCGCGGCTTATGACACATGCGAGAAAGTCTCTAC 660
 Qy 661 ATGAGAACGTTGTTATATGCGAGCTCGGCTACCGCTTCAAGGAGGCTTATG 720
 Db 661 ATGAGAACGTTGTTATATGCGAGCTCGGCTACCGCTTCAAGGAGGCTTATG 720
 Qy 721 GCTGTGAAGGACTGCGGATTCAGGGGGGGGACCTCCGTTCCATTCCCTGAGTTGGAC 780
 Db 721 GCTGTGAAGGACTGCGGATTCAGGGGGGGGACCTCCGTTCCATTCCCTGAGTTGGAC 780
 Qy 781 TTGATCAAGAGAGCCCAAGAGGCGGCGCATGAGAGGCTGGCGCGCGGCTCTACTCC 840
 Db 781 TTGATCAAGAGAGCCCAAGAGGCGGCGCATGAGAGGCTGGCGCGCGGCTCTACTCC 840
 Qy 841 AGAGCGCGCGCTCATTTTCTCCGCGAGCAGCAGCTGGGGTGAATGCTTACACGACTTC 900
 Db 841 AGAGCGCGCGCTCATTTTCTCCGCGAGCAGCAGCTGGGGTGAATGCTTACACGACTTC 900
 Qy 901 CACACATGCTGCGCTTACCTACTGCTGAGAGAACTACCCACTCCGTTGGTTTCTACTC 960
 Db 901 CACACATGCTGCGCTTACCTACTGCTGAGAGAACTACCCACTCCGTTGGTTTCTACTC 960
 Qy 961 CATATGAAGTCCAGAGGCGGCGCAGGCTCCAGAGTTAGCTGGCGCATCATCGCGCGCCA 1020
 Db 961 CATATGAAGTCCAGAGGCGGCGCAGGCTCCAGAGTTAGCTGGCGCATCATCGCGCGCCA 1020
 Qy 1021 CTTGAGACCGGATGCTGCGCGCTGCTCCAGAGAGCTCCAGAGGCTCTGAGAAATCCA 1080
 Db 1021 CTTGAGACCGGATGCTGCGCGCTGCTCCAGAGAGCTCCAGAGGCTCTGAGAAATCCA 1080
 Qy 1081 AACATCTTTTATTAAGCGCGAGCGGCTGGGCACTGAGATTAAGACATCTTACGTCAGAG 1140
 Db 1081 AACATCTTTTATTAAGCGCGAGCGGCTGGGCACTGAGATTAAGACATCTTACGTCAGAG 1140
 Qy 1141 CACAAAGTCAACATGCTGCTCAATTTTATGCAAAAGCGTCTGAGAAAGAACTAGTGGAC 1200
 Db 1141 CACAAAGTCAACATGCTGCTCAATTTTATGCAAAAGCGTCTGAGAAAGAACTAGTGGAC 1200
 Qy 1201 ATGACCGTCTTGAATGCTGCTGCGCGGCTAGGCTGGCGCGATTCGCTCAGAA 1260
 Db 1201 ATGACCGTCTTGAATGCTGCTGCGCGGCTAGGCTGGCGCGATTCGCTCAGAA 1260
 Qy 1261 GTACCGGAGGCTCTGGAACATCTCACCGGCTGACCTGCTCGGCACTGCTGATCTT 1320
 Db 1261 GTACCGGAGGCTCTGGAACATCTCACCGGCTGACCTGCTCGGCACTGCTGATCTT 1320
 Qy 1321 TAGCTCTTCAACATCACTGAGGCGCGGAGGCGTGGGCTGCTGCAAGACCGCCAAAGAC 1380
 Db 1321 TAGCTCTTCAACATCACTGAGGCGCGGAGGCGTGGGCTGCTGCAAGACCGCCAAAGAC 1380
 Qy 1381 GACGCTTTTGAAGAGACTTCTGTTCTGAGAGAGTCTCAAGACGCTTGGGCTCCTGTTGC 1440
 Db 1381 GACGCTTTTGAAGAGACTTCTGTTCTGAGAGAGTCTCAAGACGCTTGGGCTCCTGTTGC 1440
 Qy 1441 GGTCTGCGCACTTCAAGAGGACCGAGCTGCTCTTAAAGCGCTGCTGATAGTAC 1500
 Db 1441 GGTCTGCGCACTTCAAGAGGACCGAGCTGCTCTTAAAGCGCTGCTGATAGTAC 1500
 Qy 1501 CGAGTCACTCGCTCGGAGCATTAATCTGAGAGTCCGCTGCTCCCTGAAAGAGTGGC 1560
 Db 1501 CGAGTCACTCGCTCGGAGCATTAATCTGAGAGTCCGCTGCTCCCTGAAAGAGTGGC 1560

```

Db 1501 CGAAGTCACGTGCTGGAGACATTAATCTGCGACGTCGCCCTGTCCTCCCTGAAACAGGTCCGC 1560
Qy 1551 TTCTGTCGCTCCGCGTACACCTGCGCGGCTCTTTCAAGACAGGAAAGCTTGAAGTGC 1620
Db 1561 TTCTGTCGCTCCGCGTACACCTGCGCGGCTCTTTCAAGACAGGAAAGCTTGAAGTGC 1620
Qy 1621 CTTTCCGAAAGCTGCTGCTCAACAACCAACGTCCGGTAACTTTCACCTCTCTCTGTGAAGAG 1680
Db 1621 CTTTCCGAAAGCTGCTGCTCAACAACCAACGTCCGGTAACTTTCACCTCTCTGTGAAGAG 1680
Qy 1681 CCCCAAGGTTTGAAGCGCGACCTTGTGACGCGGACCGCGGCGCTCACTCCCGAGTACCGC 1740
Db 1681 CCCCAAGGTTTGAAGCGCGACCTTGTGACGCGGACCGCGGCGCTCACTCCCGAGTACCGC 1740
Qy 1741 GCCACCTTGAAGACAGTCTCAACACCGACGTCAAGACAGCTCAAGATCAACCTTCAGAAC 1800
Db 1741 GCCACCTTGAAGACAGTCTCAACACCGACGTCAAGACAGCTCAAGATCAACCTTCAGAAC 1800
Qy 1801 GCCCTCAAGACATGACGCGGCTCAACCTCTCTCCAGTGAAGGCTCGAGATGTAACGAG 1860
Db 1801 GCCCTCAAGACATGACGCGGCTCAACCTCTCTCCAGTGAAGGCTCGAGATGTAACGAG 1860
Qy 1861 GGCCTGCAAGACATGACGCGGCTCAACCTCTCTCCAGTGAAGGCTCGAGATGTAACGAG 1860
Db 1861 GGCCTGCAAGACATGACGCGGCTCAACCTCTCTCCAGTGAAGGCTCGAGATGTAACGAG 1860
Qy 1921 AAAGACCTTAAAGTGAACCAACCAAGACAGTGAAGAGGCTTGAACCGCGAGATCAAA 1980
Db 1921 AAAGACCTTAAAGTGAACCAACCAAGACAGTGAAGAGGCTTGAACCGCGAGATCAAA 1980
Qy 1981 CCGCGCTCGGCTCGGCTACGCAACATGTCGCTTGGCATTTCTCGTGTGTCACCGGC 2040
Db 1981 CCGCGCTCGGCTCGGCTACGCAACATGTCGCTTGGCATTTCTCGTGTGTCACCGGC 2040
Qy 2041 GAGGGGCGCCCTTGTGCTACCGTGAATGACGAGTGTCTTACGAGTGTCTTCCGCTCGTAC 2100
Db 2041 GAGGGGCGCCCTTGTGCTACCGTGAATGACGAGTGTCTTACGAGTGTCTTCCGCTCGTAC 2100
Qy 2101 GTGCGATGTCGACGCTTGTGTCGCGAGCTCAACGATAGTCTTGTGAAGGACGTCAC 2160
Db 2101 GTGCGATGTCGACGCTTGTGTCGCGAGCTCAACGATAGTCTTGTGAAGGACGTCAC 2160
Qy 2161 CAAATCGGTTATAGACTTCAAGGACAAAGGCGCAACATGCGCTGTCGCGACGTC 2220
Db 2161 CAAATCGGTTATAGACTTCAAGGACAAAGGCGCAACATGCGCTGTCGCGACGTC 2220
Qy 2221 GTTAAAGCAATGCGCTGCGGCACTTTCACCAACCAAGCGCTGTCGCGGCGACGTCGT 2280
Db 2221 GTTAAAGCAATGCGCTGCGGCACTTTCACCAACCAAGCGCTGTCGCGGCGACGTCGT 2280
Qy 2281 GGCACACGTTTTTTCAGAGCTTGTACCCGCGGTGCAACAACCTCAAGGTGCGTCGA 2340
Db 2281 GGCACACGTTTTTTCAGAGCTTGTACCCGCGGTGCAACAACCTCAAGGTGCGTCGA 2340
Qy 2341 TCCATGACGACGTCGCCCCAAGCTACCGCAACAGGCGGCAACGCTGTCGCAAG 2400
Db 2341 TCCATGACGACGTCGCCCCAAGCTACCGCAACAGGCGGCAACGCTGTCGCAAG 2400
Qy 2401 CAGGAGGAAAGTCCGCGCAACGCGGCTGAGGCGGCGATGACTGTGACAAAGGCGAGGA 2460
Db 2401 CAGGAGGAAAGTCCGCGCAACGCGGCTGAGGCGGCGATGACTGTGACAAAGGCGAGGA 2460
Qy 2461 CCGACTTTTGGCTGTGCTATTCATTCACCGGCTCAACAGACAGAGAGAGAGTCTCTC 2520
Db 2461 CCGACTTTTGGCTGTGCTATTCATTCACCGGCTCAACAGAGAGAGAGAGTCTCTC 2520
Qy 2521 GCTGAGAAAGTCCGCTTGAAGTCCGACATCAAGCGGCTCAACAGAGAGAGAGAGTCTCTC 2580
Db 2521 GCTGAGAAAGTCCGCTTGAAGTCCGACATCAAGCGGCTCAACAGAGAGAGAGAGTCTCTC 2580
Qy 2581 GACCGCAAGTGAACATGAGAGACAACTCAACCATAGCGCGAAGGCGAGGCTTAC 2640
Db 2581 GACCGCAAGTGAACATGAGAGACAACTCAACCATAGCGCGAAGGCGAGGCTTAC 2640

```

```

Qy 2641 GACATCCCTGACCCCTGGAATCAACGATCTCAACCGAATGAAGAGTCAAGCGCAAC 2700
Db 2641 GACATCCCTGACCCCTGGAATCAACGATCTCAACCGAATGAAGAGTCAAGCGCAAC 2700
Qy 2701 GAAATGATGGAAGATACCCCGAGAGTGCACAGCGGCGAGGAGCAATCATCTGCTC 2760
Db 2701 GAAATGATGGAAGATACCCCGAGAGTGCACAGCGGCGAGGAGCAATCATCTGCTC 2760
Qy 2761 CGCAAGAACTTCGAGGACCAACCGGCTGTGCTGTGCTGCTTGGCGAAGACCGGCTAC 2820
Db 2761 CGCAAGAACTTCGAGGACCAACCGGCTGTGCTGTGCTGCTTGGCGAAGACCGGCTAC 2820
Qy 2821 GAGGTGTTGGCGGCTCGTCCAAATCAAGTGAAGTTCGCAACCGGCGAGCGGCTAC 2880
Db 2821 GAGGTGTTGGCGGCTCGTCCAAATCAAGTGAAGTTCGCAACCGGCGAGCGGCTAC 2880
Qy 2881 AAGCGCATAGGAGGCTTCAAGAAAGGAGTCAAGTGTCAAGTCAACGAGCGCTAAC 2940
Db 2881 AAGCGCATAGGAGGCTTCAAGAAAGGAGTCAAGTGTCAAGTCAACGAGCGCTAAC 2940
Qy 2941 AAACACAGGCGCTTCAGACGCTGTGTCCGCTAACACCAAGGAGGCTGACCTGCGC 3000
Db 2941 AAACACAGGCGCTTCAGACGCTGTGTCCGCTAACACCAAGGAGGCTGACCTGCGC 3000
Qy 3001 CTACACGAAAGTAAAGAGAGAGTCAACGATGCTAACTGCTTGAACCGGCTGAG 3060
Db 3001 CTACACGAAAGTAAAGAGAGAGTCAACGATGCTAACTGCTTGAACCGGCTGAG 3060
Qy 3061 TGAGCTGTCACTGAAGAGCGCGGTGACCGAGCTGTTCGAGACCGGCTAAGTTCAC 3120
Db 3061 TGAGCTGTCACTGAAGAGCGCGGTGACCGAGCTGTTCGAGACCGGCTAAGTTCAC 3120
Qy 3121 CAACGCGGCGGACCGTGAAGACCTGTGAGGACGACGACCCCTACATCGTGAACATA 3180
Db 3121 CAACGCGGCGGACCGTGAAGACCTGTGAGGACGACGACCCCTACATCGTGAACATA 3180
Qy 3181 GACTTCTTATGAAGACTCAAGAGAGTGTGCGCCAAAGCGGATCAATACGCGGACGTC 3240
Db 3181 GACTTCTTATGAAGACTCAAGAGAGTGTGCGCCAAAGCGGATCAATACGCGGACGTC 3240
Qy 3241 GGGGAGGAGATCGCGCTCACTCAAGTCTCAACTTGTGCTGCGGCTTGTGATAGC 3300
Db 3241 GGGGAGGAGATCGCGCTCACTCAAGTCTCAACTTGTGCTGCGGCTTGTGATAGC 3300
Qy 3301 ATACTGAGAGATCTCCGTAACCGGAGCGGACGAGTCCGGTACAGCAAGTCTCCG 3360
Db 3301 ATACTGAGAGATCTCCGTAACCGGAGCGGACGAGTCCGGTACAGCAAGTCTCCG 3360
Qy 3361 GACGAAAGAGGCGATCTGCTCGAAGAGATCAATCAAGTCCCAACGCGACGTTCC 3420
Db 3361 GACGAAAGAGGCGATCTGCTCGAAGAGATCAATCAAGTCCCAACGCGACGTTCC 3420
Qy 3421 GTCTGCGGAGCTGGAACGAGTGTGAACCGCGCAATTAACAGAGAGAGTGTCTTC 3480
Db 3421 GTCTGCGGAGCTGGAACGAGTGTGAACCGCGCAATTAACAGAGAGAGTGTCTTC 3480
Qy 3481 GCGGCGCTTTTGAAGCGATCGGACAGCTGCGGTTAATCTATTCAAGAAAGG 3540
Db 3481 GCGGCGCTTTTGAAGCGATCGGACAGCTGCGGTTAATCTATTCAAGAAAGG 3540
Qy 3541 TGTGGAAGACGACCTTTCAGAGGAGGAGGCTCGTTGAAGTGAAGTGTGCTC 3600
Db 3541 TGTGGAAGACGACCTTTCAGAGGAGGAGGCTCGTTGAAGTGAAGTGTGCTC 3600
Qy 3601 GACTCGGCGAGCTTGAAGCGCTTGGCGCAACCAATCTTCTGCGCGCTCAATGCTC 3660
Db 3601 GACTCGGCGAGCTTGAAGCGCTTGGCGCAACCAATCTTCTGCGCGCTCAATGCTC 3660
Qy 3661 ACGCTCTTCGCGGAGTCAAGTTCGACACTTTCAAGGAGAGAGTCTCGCTCTGTGCT 3720
Db 3661 ACGCTCTTCGCGGAGTCAAGTTCGACACTTTCAAGGAGAGAGTCTCGCTCTGTGCT 3720

```

QY 3721 AGCCATTACTCGGTTTCGACGCTAGCCGCTTCACTAGGCGGAAGCTTACAAAGCCAAA 3780
 DB 3721 AGCCATTACTCGGTTTCGACGCTAGCCGCTTCACTAGGCGGAAGCTTACAAAGCCAAA 3780
 QY 3781 CATTGGAAGGTCGAGGTCAGAAATGTCGCGTACATCGGACCTCGTCCGCTGAG 3840
 DB 3781 CATTGGAAGGTCGAGGTCAGAAATGTCGCGTACATCGGACCTCGTCCGCTGAG 3840
 QY 3841 CAGTGTCTCGACCCCTGTCAAGAGCGCTCTCAAGATATTTGGCGCTGTCAACAAGC 3900
 DB 3841 CAGTGTCTCGACCCCTGTCAAGAGCGCTCTCAAGATATTTGGCGCTGTCAACAAGC 3900
 QY 3901 GAATCTCTTACTTCCAAAGTACGTGAGAGCTGTGAGAGATCAACAAAGGCTGAGTAC 3960
 DB 3901 GAATCTCTTACTTCCAAAGTACGTGAGAGCTGTGAGAGATCAACAAAGGCTGAGTAC 3960
 QY 3961 GCCCGCTACAGAGCCCTCTGTGCGACATGTCAAGATGCTACTACAAATTAACGCGCGAG 4020
 DB 3961 GCCCGCTACAGAGCCCTCTGTGCGACATGTCAAGATGCTACTACAAATTAACGCGCGAG 4020
 QY 4021 TCTGCGCGTACATCATCGACGCTGTGTGTTGCTTTGGGCGCGGCGACTTCCGTTTGA 4080
 DB 4021 TCTGCGCGTACATCATCGACGCTGTGTGTTGCTTTGGGCGCGGCGACTTCCGTTTGA 4080
 QY 4081 CAATCGCGCTGTGCTGCTGCTCATGTGAGAGCACCCGACTTACAGACAGATATCCG 4140
 DB 4081 CAATCGCGCTGTGCTGCTGCTCATGTGAGAGCACCCGACTTACAGACAGATATCCG 4140
 QY 4141 GCTAAGCGTGGCGATGTGCTGCTTGAACAGTCTTGAAGCCGCGCGAGCGCGCGCCG 4200
 DB 4141 GCTAAGCGTGGCGATGTGCTGCTTGAACAGTCTTGAAGCCGCGCGAGCGCGCGCCG 4200
 QY 4201 GCAGGTTTCTGTGCAATGTGCGAAGCCGGAAGCGCTTCTTCACTTACCGGAAAGCT 4260
 DB 4201 GCAGGTTTCTGTGCAATGTGCGAAGCCGGAAGCGCTTCTTCACTTACCGGAAAGCT 4260
 QY 4261 GGTGTTTCTGCGACTACAGACGCTTGGCACTGCGGCTGCGGAGCTTCCATGCG 4320
 DB 4261 GGTGTTTCTGCGACTACAGACGCTTGGCACTGCGGCTGCGGAGCTTCCATGCG 4320
 QY 4321 GATGCACTGTGAGCAAGCAAGCTTTGGAGTATTTGACACCGGAGACCCGCTCCACATCA 4380
 DB 4321 GATGCACTGTGAGCAAGCAAGCTTTGGAGTATTTGACACCGGAGACCCGCTCCACATCA 4380
 QY 4381 TCTCGCGCTGATGCTTTCATCGGACTCCTCTCATCGTGTGGAAGTGCCTCATGCT 4440
 DB 4381 TCTCGCGCTGATGCTTTCATCGGACTCCTCTCATCGTGTGGAAGTGCCTCATGCT 4440
 QY 4441 GGAGACACCGCGAAGACCAAGAAAGCTTGAACAGCAAGAAAGCCGCTTGGCAAGACAG 4500
 DB 4441 GGAGACACCGCGAAGACCAAGAAAGCTTGAACAGCAAGAAAGCCGCTTGGCAAGACAG 4500
 QY 4501 CAATCAAGCTGTGTAATGTCTGGAAGAGCGGAGGAGGAGGAGGAGTTCCTTAAT 4560
 DB 4501 CAATCAAGCTGTGTAATGTCTGGAAGAGCGGAGGAGGAGGAGGAGTTCCTTAAT 4560
 QY 4561 GCCCGCACTGCTCGAGACCCCTCATTTCTCATTTTTCGAAAGAGTTCGACTGCGACCGG 4620
 DB 4561 GCCCGCACTGCTCGAGACCCCTCATTTCTCATTTTTCGAAAGAGTTCGACTGCGACCGG 4620
 QY 4621 CCGACTGTGCGCGTGTGAGCATCACTTCCGCAACCCCATCTGCGCACGAGCAAGGTT 4680
 DB 4621 CCGACTGTGCGCGTGTGAGCATCACTTCCGCAACCCCATCTGCGCACGAGCAAGGTT 4680
 QY 4681 GCGCGAGAGACCAAGCGGAGCTTTGCGCTTCTGAGTTCCTGAGTTCGCGGCTGCTGTC 4740
 DB 4681 GCGCGAGAGACCAAGCGGAGCTTTGCGCTTCTGAGTTCCTGAGTTCGCGGCTGCTGTC 4740
 QY 4741 TCGAAGCGTACCGGCGCCCGACGACTGCGCTTGGAAAGAGTCAACCCGCTCCACGCG 4800
 DB 4741 TCGAAGCGTACCGGCGCCCGACGACTGCGCTTGGAAAGAGTCAACCCGCTCCACGCG 4800
 QY 4801 TGGAAAGGCGTGAACCGGAGACCGAGAGTCAAGGAGAGACCCGAGAGACGCGCGGTC 4860

DB 4801 TGGAAAGGCGTGAACCGGAGACCGAGAGTCAAGGAGAGACCCGAGAGACGCGCGGTC 4860
 QY 4861 GTTCAGGCTCTGATCAAGCGCGCTTATCTCTCAGAAAGCAAGCTTTCCTCCAGCATCC 4920
 DB 4861 GTTCAGGCTCTGATCAAGCGCGCTTATCTCTCAGAAAGCAAGCTTTCCTCCAGCATCC 4920
 QY 4921 AAAGGCTACTCAAGAACTAAAGGATGCTCAACAATCACTCTTTTCTCTGCGCGAGTCCG 4980
 DB 4921 AAAGGCTACTCAAGAACTAAAGGATGCTCAACAATCACTCTTTTCTCTGCGCGAGTCCG 4980
 QY 4981 GATTACAGGCGCGAGCTGCGAGAGTCCGAGTCTGCGCGCGCGCTGCGAGAGTGGCG 5040
 DB 4981 GATTACAGGCGCGAGCTGCGAGAGTCCGAGTCTGCGCGCGCGCTGCGAGAGTGGCG 5040
 QY 5041 CGCTCATGTATTCACAGAGCGGTTGCTTCAATCTGCGCGAGTCCGAGCTTGAAGCGCAT 5100
 DB 5041 CGCTCATGTATTCACAGAGCGGTTGCTTCAATCTGCGCGAGTCCGAGCTTGAAGCGCAT 5100
 QY 5101 CGCTCACTCGGACTCTGTGTTCCCATGTAAAGATCAAGCAAGCGCATGAAGCAAA 5160
 DB 5101 CGCTCACTCGGACTCTGTGTTCCCATGTAAAGATCAAGCAAGCGCATGAAGCAAA 5160
 QY 5161 ATTAGTTTCTGTTGTGTAACAAAGGAGTCCCTCCATGAGGTAAAGACTCTGAGTAC 5220
 DB 5161 ATTAGTTTCTGTTGTGTAACAAAGGAGTCCCTCCATGAGGTAAAGACTCTGAGTAC 5220
 QY 5221 TCTCAAGCTTACTCTGTGATGCTGCTGCGGTTGATTCATTCAGAGCAAGAGGCT 5280
 DB 5221 TCTCAAGCTTACTCTGTGATGCTGCTGCGGTTGATTCATTCAGAGCAAGAGGCT 5280
 QY 5281 GCGCAACTAGTACGCGCGCCCTGCGATACCA 5312
 DB 5281 GCGCAACTAGTACGCGCGCCCTGCGATACCA 5312
 RESULT 4
 HAU18246
 LOCUS
 DEFINITION
 Helicoverpa armigera stunt virus RNA-dependent RNA polymerase, complete cds.
 VERSION
 U18246.1 GI:603222
 KEYWORDS
 SOURCE
 ORGANISM
 Helicoverpa armigera stunt virus
 Helicoverpa armigera stunt virus
 Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae; unclassified Tetraviridae.
 REFERENCE
 1 (bases 1 to 5312)
 Gordon, K.H., Johnson, K.N. and Hanzlik, T.N.
 The larger genomic RNA of Helicoverpa armigera stunt tetravirus encodes the viral RNA polymerase and has a novel 3'-terminal tRNA-like structure
 JOURNAL
 virology 208 (1), 84-98 (1995)
 MEDLINE
 21820274
 PUBMED
 11831734
 REFERENCE
 2 (bases 1 to 5312)
 Gordon, K.H.J.
 Direct Submission
 Submitted (07-DEC-1994) Karl H.J. Gordon, CSIRO Division of Entomology, Biotechnology Section, Canberra, Act, Australia, 2601
 FEATURES
 Location/Qualifiers
 1..5312
 /organism="Helicoverpa armigera stunt virus"
 /mol_type="genomic RNA"
 /specific_host="Helicoverpa armigera"
 /db_xref="taxon:37206"
 37..5151
 /codon_start=1
 /product="RNA-dependent RNA polymerase"
 /protein_id="AAC98529.1"
 /db_xref="gi:603222"
 /translation="MTAKRDVAVRYAADVAIVANVLQQRADVLDPAFPLKALETLHR


```

LYPLRFKGTLLPPTPHIAGHOREVAVELNFAGRSTVLEIGPSHSLKILHGAP
NAPVADHYGCTKYGRDSDSRHTTALLESBSVATGRPEPRADSLANGIASRPTCDGV
GSCAFKSVGIANHSLYDTLELANAFENGLHNVRAFMHPEELLTMDVNVNELG
YRFVIEBPMAVKDCAFQGGDLRLHFPELDFINSEOBRIELRLAGVSRAVIFSG
DDMDGADYLDHPTMLAVLYVNSYPTPGFSLIHVEORHSGSILRTYRAIPGRML
AVPRTSGDLGRIPIPIFYADASGTEHKTILTSOKHMLNFMOTREKELVDVTL
MSPARLRALIVVASEVTESSWNISPADLVTVSVLYLIIERRBAVAVTADQV
FGEISFMSLKHVLSSCCGLRNKGTDFVTRVVDKTRVHSLGDIICDVRLSPQV
FLPSRVPARVPHDELELEVELEACVNERPVSTPVEBPQGFADMLHNTASLPE
YRATLQAGLNDVKQKLTLEENALKTIDGLTISPVRGLEMEGAPSGSTGLTIALB
AAGKALVYAPPRELEBAMDRRIKPPSASRTQHVALLALIRAGSAPATVYIDEC
MEPLVYVIVHALSPSSRTVLVGDVHOGIFIDPGTSAMPVLCVPOEBCRRTNOT
KCPADVATTPEOSLPGCTTSGCVASISVADPNYSQAQCTCFPOEBSRBAK
GAMTHEOGKTRFASVILIHNGSINABQKILAKSHULVIGTRHNTHLIIRDTGIER
QUNHSAKAEVFTDIPAPLEITTVKPSSEVQNEVMATIPPOSPTHGA1HLIKRPF
QPDCCVALAKTGYEFGGRAKINVELPEPMPKPHAPPOGVOMVYTNASNGHA
LQTLISRYTKRSADLPLEHAKEDVKNMNSLDHMDWTGEBARAPRAVEITLKTOR
GGTVEDLLEDDPYIRIDIDFLMKTOOKVSPKPIINTGKQGGIHAASKSLNFIAAMIR
LLEILRTGSRTRVRSNGLPDEEAMLEAKINOVPHATGVAGIADMTREDDTANNSTEL
LEPALLERICTPAAAVNPREFCGKRTLRKAGLSVEVDGLDSGAATPCANTTFS
AVMLTLFPGVKFAFRKGDLSLCSHYLEPDSRLHMEGRYTKELKVEVQKIVYIG
LIVSAEOVVLDEVRSAIKIFGRCTSELSTYSKYVAHVODITGMSDARVHSLCHMSA
CYNYAPESAAVYIDAVVFRGDPPEQLRVRAHVODPDAVSSYTPANVASCIDH
VEEPROAAPAFVATCAKPEPSSILTAAGVSATTSVATSTAPESPMDAPANSF
SELTTPBPSTSSSAVIVRIGLYIVMKVAMWMBRKRKREDEINSRKRPSODQSSSE
CLDRSGERTGSLTAPLAPSPSFSERKRLTGPVAAISPSATPSCATDQVART
TPDFAPFLGOSARAVSKRPPTTARMEVTPPLAHAMKVOTDRREVREDPTAAVVO
ALISGRYPOKTLSSDASKVSGRTKGSSTFPAPSPADYQARDQOTVAVCRBAAMEA
SICHPPLASSASADLKRIRSTSDSPVPIKSKSA"
5190..5312
misc_structure
/note="3'-terminal tRNA-like structure can be formed by
this region; lacks pseudo-knot in aminoacyl acceptor stem;
has valine anticodon"
BASE COUNT 1133 a 1739 c 1433 g 1007 t
ORIGIN
Query Match 99.4%; Score 5281; DB 14; Length 5312;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 5306; Conservative 0; Mismatches 5; Indels 2; Gaps 2;

```

```

DB 481 GGCATTGCTCCCGCACCTTCTGCGTGAAGAGTGCGGCTTGGCGGTTCAAAATGCCG 540
QY 481 GGCATTGCTCCCGCACCTTCTGCGTGAAGAGTGCGGCTTGGCGGTTCAAAATGCCG 540
DB 541 GTTGAATTCGCAATCACTCCCTCTATAGACGTGACCTTAGAGAGCTGGCCATGCGTTT 600
QY 541 GTTGAATTCGCAATCACTCCCTCTATAGACGTGACCTTAGAGAGCTGGCCATGCGTTT 600
DB 541 GTTGAATTCGCAATCACTCCCTCTATAGACGTGACCTTAGAGAGCTGGCCATGCGTTT 600
QY 601 GAGAACCGAGACTTCAATAGTCCGCGGCTTTCATGCAATGCCAGAAAGACTGCTCTAC 660
DB 601 GAGAACCGAGACTTCAATAGTCCGCGGCTTTCATGCAATGCCAGAAAGACTGCTCTAC 660
QY 601 GAGAACCGAGACTTCAATAGTCCGCGGCTTTCATGCAATGCCAGAAAGACTGCTCTAC 660
DB 601 GAGAACCGAGACTTCAATAGTCCGCGGCTTTCATGCAATGCCAGAAAGACTGCTCTAC 660
QY 661 ATGGAACAAGTGTATATAGCGAGTCCGACTACCGCTTCCAGCTATATGAAGAGCTATG 720
DB 661 ATGGAACAAGTGTATATAGCGAGTCCGACTACCGCTTCCAGCTATATGAAGAGCTATG 720
QY 721 GCTGTGAAGAGACTGCGCATTCAGAGGGGGGAGACTCCGCTTCCACTTCCTGAGTTGAC 780
DB 721 GCTGTGAAGAGACTGCGCATTCAGAGGGGGGAGACTCCGCTTCCACTTCCTGAGTTGAC 780
QY 781 TTCAATCAAGAGCCCAAGAGAGCGGCGCATTCAGAGAGCTGGCCGCGGCTCTCACTCC 840
DB 781 TTCAATCAAGAGCCCAAGAGAGCGGCGCATTCAGAGAGCTGGCCGCGGCTCTCACTCC 840
QY 841 AGACGCGCGCTCATTTTCTCCGCGAGACGACGACTGAGAGAGCTGGCCGCGGCTCTCACTCC 900
DB 841 AGACGCGCGCTCATTTTCTCCGCGAGACGACGACTGAGAGAGCTGGCCGCGGCTCTCACTCC 900
QY 901 CACACATGAGCTGCTGACTTACTGAGTGAAGAACTACCCCACTCCGTTTGTTTCACTC 960
DB 901 CACACATGAGCTGCTGACTTACTGAGTGAAGAACTACCCCACTCCGTTTGTTTCACTC 960
QY 961 CATATAGAAGTCCAGAGGCGCCAGCGCTTCCAGCATTTGAGCTGCGCATCACTGCGCGCA 1020
DB 961 CATATAGAAGTCCAGAGGCGCCAGCGCTTCCAGCATTTGAGCTGCGCATCACTGCGCGCA 1020
QY 1021 CCTGGAAGACCGGATCTGAGCGCTGCTCCCAAGAGAGCTCCCAAGAGCTTCCAGAAATCCCA 1080
DB 1021 CCTGGAAGACCGGATCTGAGCGCTGCTCCCAAGAGAGCTCCCAAGAGCTTCCAGAAATCCCA 1080
QY 1081 AACATCTTTATTAAGCGGAGCGGCTGAGGAGCTAGAGCATTAAGACCATCTTACGTCACAG 1140
DB 1081 AACATCTTTATTAAGCGGAGCGGCTGAGGAGCTAGAGCATTAAGACCATCTTACGTCACAG 1140
QY 1141 CACAAAGTCAAGTCTGCTCAATTTTATGCAAAAGCGTCTTGAGAGAACTAGTCCAG 1200
DB 1141 CACAAAGTCAAGTCTGCTCAATTTTATGCAAAAGCGTCTTGAGAGAACTAGTCCAG 1200
QY 1201 ATGACCGTCTTGAATGCTGCTGCGCGGCTGAGGCTGCGGCGGAGTGTGGTGCCTCAGAA 1260
DB 1201 ATGACCGTCTTGAATGCTGCTGCGCGGCTGAGGCTGCGGCGGAGTGTGGTGCCTCAGAA 1260
QY 1261 GTCAACGAGAGCTCTGGAACATCTCACCGCTGACCTGGTCCGACCTGTGCTCTT 1320
DB 1261 GTCAACGAGAGCTCTGGAACATCTCACCGCTGACCTGGTCCGACCTGTGCTCTT 1320
QY 1321 TACGTCCTTCAATCATGAGGCGCGGAGAGGCTGCGGTGCTGTCAAGACCGCAAGAGAC 1380
DB 1321 TACGTCCTTCAATCATGAGGCGCGGAGAGGCTGCGGTGCTGTCAAGACCGCAAGAGAC 1380
QY 1381 GACGCTTTTGAAGAGACTTGTGCTGAGAGAGTCAAGCAAGTCTTGGGCTCTGTTTGC 1440
DB 1381 GACGCTTTTGAAGAGACTTGTGCTGAGAGAGTCAAGCAAGTCTTGGGCTCTGTTTGC 1440
QY 1441 GGTCTGCGCAACTCAAAAGGACCGAGCGTGTCTTTACTAAGCGGCTGTGATAAGTAC 1500
DB 1441 GGTCTGCGCAACTCAAAAGGACCGAGCGTGTCTTTACTAAGCGGCTGTGATAAGTAC 1500
QY 1501 CGAGTCACTGCTGAGAGACATTAATCTGCAAGTCTGCTGTCTCCCTGAAACAGGTGCGC 1560
DB 1501 CGAGTCACTGCTGAGAGACATTAATCTGCAAGTCTGCTGTCTCCCTGAAACAGGTGCGC 1560
QY 1561 TTCTGCGCTCCCGGATCAACTGCGCGGCTTTCAGACAGAGGAAGAGCTTAGAGTCC 1620
DB 1561 TTCTGCGCTCCCGGATCAACTGCGCGGCTTTCAGACAGAGGAAGAGCTTAGAGTCC 1620

```

Db 1561 TTCTGCGGCTCCGCGTACCACTGCGCGCTCTTTCAACAAGAGAGGCTTGAAGTC 1620
 Qy 1621 CTTGCGAGAGCTGCTGCTCTCAACAGAAAGTCGAGTACCTTCACTCTCTGTGAGAG 1680
 Db 1621 CTTGCGAGAGCTGCTGCTCTCAACAGAAAGTCGAGTACCTTCACTCTCTGTGAGAG 1680
 Qy 1681 CCCCCAAGTTTTCAGCGCGCACTTTGTGGACGGGACCGCGGCTCACTCTCCGAGTACCGC 1740
 Db 1681 CCCCCAAGTTTTCAGCGCGCACTTTGTGGACGGGACCGCGGCTCACTCTCCGAGTACCGC 1740
 Qy 1741 GCCACTTTCAGAGAGGCTCTCAACACCGACGCTCAAGAGAGCTCAAGATCAACCTTGAAGAC 1800
 Db 1741 GCCACTTTCAGAGAGGCTCTCAACACCGACGCTCAAGAGAGCTCAAGATCAACCTTGAAGAC 1800
 Qy 1801 GCCCTCAAGACCATCGAGGGCTCACCTCTCCCAAGTCAAGAGGCTCTGAGATGACGAG 1860
 Db 1801 GCCCTCAAGACCATCGAGGGCTCACCTCTCCCAAGTCAAGAGGCTCTGAGATGACGAG 1860
 Qy 1861 GCGCCGCGACAGGACGGGCAAGAGGGGACCTCTATCGCGCGCTTGAAGCGCGGGCGGT 1920
 Db 1861 GCGCCGCGACAGGACGGGCAAGAGGGGACCTCTATCGCGCGCTTGAAGCGCGGGCGGT 1920
 Qy 1921 AAAGCACTTTACGTGGGACCCACACAGAGAACTGAGAGGCTATGAGACCGGCGGATCAA 1980
 Db 1921 AAAGCACTTTACGTGGGACCCACACAGAGAACTGAGAGGCTATGAGACCGGCGGATCAA 1980
 Qy 1981 CCGCGGCTCGGCTGAGGCTACGCAACATGTCGCGCTTGAGATTCTCCGTGTCACCGGC 2040
 Db 1981 CCGCGGCTCGGCTGAGGCTACGCAACATGTCGCGCTTGAGATTCTCCGTGTCACCGGC 2040
 Qy 2041 GAGGCGCGCTCTGCTGCTACGCTGTATGACAGAGTCTTCAAGTTTCCGCTGTGTAC 2100
 Db 2041 GAGGCGCGCTCTGCTGCTACGCTGTATGACAGAGTCTTCAAGTTTCCGCTGTGTAC 2100
 Qy 2101 GTGCGATCGTGCAGCGCTTGTCCCGAGCTCAAGAAATGCTTGTAGAGGAGAGTCCAC 2160
 Db 2101 GTGCGATCGTGCAGCGCTTGTCCCGAGCTCAAGAAATGCTTGTAGAGGAGAGTCCAC 2160
 Qy 2161 CAAATCGGGTTTATAGATTCCAAAGGACAAAGCGGAAATGCGGCTGTTCCGACGTC 2220
 Db 2161 CAAATCGGGTTTATAGATTCCAAAGGACAAAGCGGAAATGCGGCTGTTCCGACGTC 2220
 Qy 2221 GTTAAAGAGTGCCTGCGGCGCACTTTCAAACCAAGCGCTGTCCGCGAGTGGT 2280
 Db 2221 GTTAAAGAGTGCCTGCGGCGCACTTTCAAACCAAGCGCTGTCCGCGAGTGGT 2280
 Qy 2281 GCGACACGTTTTCAGAGCTTGTACCCCGGGTGCACCAACCTCAAGGGTGGTGC 2340
 Db 2281 GCGACACGTTTTCAGAGCTTGTGTACCCCGGGTGCACCAACCTCAAGGGTGGTGC 2340
 Qy 2341 TCCATCAGCCAGTGCGCCCAAGTACCGCAACAGCGAGCGCAACGCTTGTTCACG 2400
 Db 2341 TCCATCAGCCAGTGCGCCCAAGTACCGCAACAGCGAGCGCAACGCTTGTTCACG 2400
 Qy 2401 CAGAGAGAAAGTGCAGCGACAGGGGCTGAGGGCGGATGATGTGACAGAAAGCGAGGA 2460
 Db 2401 CAGAGAGAAAGTGCAGCGACAGGGGCTGAGGGCGGATGATGTGACAGAAAGCGAGGA 2460
 Qy 2461 CGCACTTTGTGCTGTGATCTGATTTCAACAGGCTTCAAGAGAGCAAGAGTCTCTC 2520
 Db 2461 CGCACTTTGTGCTGTGATCTGATTTCAACAGGCTTCAAGAGAGCAAGAGTCTCTC 2520
 Qy 2521 GCTGAGAGTGCACCTTCTAGTGGGATCAAGCGCCACACCAACCTGTATACCGC 2580
 Db 2521 GCTGAGAGTGCACCTTCTAGTGGGATCAAGCGCCACACCAACCTGTATACCGC 2580
 Qy 2581 GACCCGACAGGTGACTTGAAGACACTCAACATAGCGGAAAGCGGAGTGTTTACA 2640
 Db 2581 GACCCGACAGGTGACTTGAAGACACTCAACATAGCGGAAAGCGGAGTGTTTACA 2640
 Qy 2641 GACATCCCTGACCGCTGAGATGACGATGTTCAACCGAGTGAAGAGGTGACAGGCAAC 2700
 Db 2641 GACATCCCTGACCGCTGAGATGACGATGTTCAACCGAGTGAAGAGGTGACAGGCAAC 2700

Qy 2701 GAAATGATGGCAACGATACCCCGCAGAGTGCACAGCGCGCAAGAGCAATCATGTGCTC 2760
 Db 2701 GAAATGATGGCAACGATACCCCGCAGAGTGCACAGCGCGCAAGAGCAATCATGTGCTC 2760
 Qy 2761 CGCAGAACTTTCGGGAGCAACCGGACGTGGGCTGTGTCCCTTGTGGAGAACCGGCTAC 2820
 Db 2761 CGCAGAACTTTCGGGAGCAACCGGACGTGGGCTGTGTCCCTTGTGGAGAACCGGCTAC 2820
 Qy 2821 GAGGTGTTTGGCGGTCTGTGCAAAATCAAGTGAAGCTTGCAGAACCCGACGCGACCCG 2880
 Db 2821 GAGGTGTTTGGCGGTCTGTGCAAAATCAAGTGAAGCTTGCAGAACCCGACGCGACCCG 2880
 Qy 2881 AAGCCGATAGGGCTTTCAGAAAGGGGTACAGTGGTCAAGTACCAACGCTTAC 2940
 Db 2881 AAGCCGATAGGGCTTTCAGAAAGGGGTACAGTGGTCAAGTACCAACGCTTAC 2940
 Qy 2941 AAACACAGGCGCTCCAGAGCGCTGTGTCCCGCTACACCAAGGCAAGCGCTGACGTCG 3000
 Db 2941 AAACACAGGCGCTCCAGAGCGCTGTGTCCCGCTACACCAAGGCAAGCGCTGACGTCG 3000
 Qy 3001 CTACACGAACTAAGAGAGAGCTCAACAGCATGTAACTCGCTTGAACCGACATTGGAGAC 3060
 Db 3001 CTACACGAACTAAGAGAGAGCTCAACAGCATGTAACTCGCTTGAACCGACATTGGAGAC 3060
 Qy 3061 TGAAGTCTCACTGAAGACGCGCTGACCGAGCTGTCTTCAAGACCCAGCTCAAGTTACC 3120
 Db 3061 TGAAGTCTCACTGAAGACGCGCTGACCGAGCTGTCTTCAAGACCCAGCTCAAGTTACC 3120
 Qy 3121 CAACGCGGCGGACCGCTGAAAGACCTGTGAGAGCAACGACCCCTTCACTCCGAGACTA 3180
 Db 3121 CAACGCGGCGGACCGCTGAAAGACCTGTGAGAGCAACGACCCCTTCACTCCGAGACTA 3180
 Qy 3181 GACTTCTTATGAAGACTCAGCAGAAAGTTCGCGCAAGCGATCAATTAAGGCAAGTCC 3240
 Db 3181 GACTTCTTATGAAGACTCAGCAGAAAGTTCGCGCAAGCGATCAATTAAGGCAAGTCC 3240
 Qy 3241 GGGGAGGGAGTCCGCGCTCACTCAAGTCTCTCAACTTCGCTCCGCGCTTGAATACG 3300
 Db 3241 GGGGAGGGAGTCCGCGCTCACTCAAGTCTCTCAACTTCGCTCCGCGCTTGAATACG 3300
 Qy 3301 ATACTGAGAGAGTATCTCCGTAACCGGAGGCGCAAGGTCGCTCCGCTCCGCTCCG 3360
 Db 3301 ATACTGAGAGAGTATCTCCGTAACCGGAGGCGCAAGGTCGCTCCGCTCCGCTCCG 3360
 Qy 3361 GACGAAAGAGGCGCATGTGCTGCAAGAGAAAGTCAATCAATGCGCACAGCGCACTTC 3420
 Db 3361 GACGAAAGAGGCGCATGTGCTGCAAGAGAAAGTCAATCAATGCGCACAGCGCACTTC 3420
 Qy 3421 GTCTGCGGAGCTGAGACGAGTTTGAACAGCGGCCCAATTAACAGATGAGCTGCTTC 3480
 Db 3421 GTCTGCGGAGCTGAGACGAGTTTGAACAGCGGCCCAATTAACAGATGAGCTGCTTC 3480
 Qy 3481 GCGCGCTTTTAAAGCGCATTCGCAAGCGCTGCAAGCTCCGTTAATCTAATCAAGAAAG 3540
 Db 3481 GCGCGCTTTTAAAGCGCATTCGCAAGCGCTGCAAGCTCCGTTAATCTAATCAAGAAAG 3540
 Qy 3541 TGTGGAGAAAGCACTTTCGAGAGCAAGGGTCTAGGCTTCGTTGAAGTGCAGGCTGCTC 3600
 Db 3541 TGTGGAGAAAGCACTTTCGAGAGCAAGGGTCTAGGCTTCGTTGAAGTGCAGGCTGCTC 3600
 Qy 3601 GACTCGGCGGAGTGAAGCGCTTTCGCGCAACCAATCTTCTGCGCGCGGTCAATGCTC 3660
 Db 3601 GACTCGGCGGAGTGAAGCGCTTTCGCGCAACCAATCTTCTGCGCGCGGTCAATGCTC 3660
 Qy 3661 ACGTCTTTCGCGCGGCTCAAGTTTTCAGAGCTTTCAAAGCGCAAGCTGCTCTGTGGT 3720
 Db 3661 ACGTCTTTCGCGCGGCTCAAGTTTTCAGAGCTTTCAAAGCGCAAGCTGCTCTGTGGT 3720
 Qy 3721 AGCCATTACCTCGTTTGAAGCTAGCGGCTTCAATGAGCGAAAGTTTCAAGACCAA 3780
 Db 3721 AGCCATTACCTCGTTTGAAGCTAGCGGCTTCAATGAGCGAAAGTTTCAAGACCAA 3780

QY 3781 CATTGAAGGTGAGGTGAGAAATCGTCCGTACATCGACTCTCGTCCGCTAG 3840
 DB 3781 CATTGAAGGTGAGGTGAGAAATCGTCCGTACATCGACTCTCGTCCGCTAG 3840
 QY 3841 CAGGTGCTGAGACCTGTGAGAGCGCTCTCAAGATATTTGGCGCTGTACACAAG 3900
 DB 3841 CAGGTGCTGAGACCTGTGAGAGCGCTCTCAAGATATTTGGCGCTGTACACAAG 3900
 QY 3901 GAATCTCTTTAATCTCAAGTACGTGAGAGCGCTGTAGAGACATCAACAAGGCTGAGTAC 3960
 DB 3901 GAATCTCTTTAATCTCAAGTACGTGAGAGCGCTGTAGAGACATCAACAAGGCTGAGTAC 3960
 QY 3961 GCGGCTTACCAAGCGCTCTGTGAGACATGTGAGATGCTACTAATAATGCGCGGAG 4020
 DB 3961 GCGGCTTACCAAGCGCTCTGTGAGACATGTGAGATGCTACTAATAATGCGCGGAG 4020
 QY 4021 TCTGCGGCTACATCATGACGCTTTGCTTTGGCGCGGAGCTTCCGTTTGA 4080
 DB 4021 TCTGCGGCTACATCATGACGCTTTGCTTTGGCGCGGAGCTTCCGTTTGA 4080
 QY 4081 CAATGCGGTGAGTGTGAGAGCGCTGTGAGAGCGCTGTACAGAGCGTATCCG 4140
 DB 4081 CAATGCGGTGAGTGTGAGAGCGCTGTGAGAGCGCTGTACAGAGCGTATCCG 4140
 QY 4141 GCTAACGTGCGGCTGCTGCTTTGAGACGCTTTGAGCGCGCGCGCGCGCG 4200
 DB 4141 GCTAACGTGCGGCTGCTGCTTTGAGACGCTTTGAGCGCGCGCGCGCGCGCG 4200
 QY 4201 GCAGGTTTCTGTTGAGACATGTGAGAGCGCGAAACGCTTTCTTCACTTACCGGAAAGT 4260
 DB 4201 GCAGGTTTCTGTTGAGACATGTGAGAGCGCGAAACGCTTTCTTCACTTACCGGAAAGT 4260
 QY 4261 GGTGTTTCTGAGCTACCAAGCGTGTGAGCTGCGCGCGCGCGAGTCTCCATG 4320
 DB 4261 GGTGTTTCTGAGCTACCAAGCGTGTGAGCTGCGCGCGCGCGAGTCTCCATG 4320
 QY 4321 GATGACCTGAGCGCAACGCTTTTCTGAGTATGACCGGAGAGCGCGCGCGCATCA 4380
 DB 4321 GATGACCTGAGCGCAACGCTTTTCTGAGTATGACCGGAGAGCGCGCGCGCATCA 4380
 QY 4381 TCCCTC-GCCGTCATGCTTCTCATCGGACTCTCTACATGCTGTGAGAGTCTGCTAGTGG 4439
 DB 4381 TCCCTC-GCCGTCATGCTTCTCATCGGACTCTCTACATGCTGTGAGAGTCTGCTAGTGG 4439
 QY 4440 TGGAGACACCGGAGAGCAAGAGCTTTGAAACAGCAGAGAGCGCTTCCGAGACAG 4499
 DB 4440 TGGAGACACCGGAGAGCAAGAGCTTTGAAACAGCAGAGAGCGCTTCCGAGACAG 4499
 QY 4441 TGGAGACACCGGAGAGCAAGAGCTTTGAAACAGCAGAGAGCGCTTCCGAGACAG 4499
 DB 4441 TGGAGACACCGGAGAGCAAGAGCTTTGAAACAGCAGAGAGCGCTTCCGAGACAG 4499
 QY 4500 GCAATTCAGCTGCTGTGATGTCTGAGACAGAGCGGAGAGAGGAGGAGTCTTAA 4559
 DB 4500 GCAATTCAGCTGCTGTGATGTCTGAGACAGAGCGGAGAGAGGAGGAGTCTTAA 4559
 QY 4560 TGCCTCCATGCTGCTGAGAGCGCTTCTCATTTTGGAGAGAGCTGAGTGGACCG 4619
 DB 4560 TGCCTCCATGCTGCTGAGAGCGCTTCTCATTTTGGAGAGAGCTGAGTGGACCG 4619
 QY 4620 GCGGACTGTGCGCTGAGACATCACTTCCGCAACCCCATCTGTGCGGACGAGT 4679
 DB 4620 GCGGACTGTGCGCTGAGACATCACTTCCGCAACCCCATCTGTGCGGACGAGT 4679
 QY 4680 TGGCGGAGAGACACCGCGGAGCTTTGCGGCTTCCGAGTCTGCGGAGTCT 4739
 DB 4680 TGGCGGAGAGACACCGCGGAGCTTTGCGGCTTCCGAGTCTGCGGAGTCT 4739
 QY 4740 CTGGAAGCGTACCGGCGCGGAGCTGCGGCTTGGAGAGAGTCAACCGCTCCACG 4799
 DB 4740 CTGGAAGCGTACCGGCGCGGAGCTGCGGCTTGGAGAGAGTCAACCGCTCCACG 4799
 QY 4800 GTGGAAGGCGGTGACCGGAGACCGAGCGGAGTCAAGGAGAGACCGGAGACAGCGCGGT 4859
 DB 4800 GTGGAAGGCGGTGACCGGAGACCGAGCGGAGTCAAGGAGAGACCGGAGACAGCGCGGT 4859
 QY 4860 CGTCCAGGCTGTGATGAGCGCGGCTTATCTTCAAGAGAGAGCTTCTCCGAGCGCATC 4919

DB 4860 CGTCCAGGCTGTGATGAGCGCGGCTTATCTTCAAGAGAGAGCTTCTCCAGCGATC 4919
 QY 4920 CAAGGCTACTCAAGAACTAAGGATGCTCAATTCACCTCTTTTCTGCGCGAGTGC 4979
 DB 4920 CAAGGCTACTCAAGAACTAAGGATGCTCAATTCACCTCTTTTCTGCGCGAGTGC 4979
 QY 4980 GGAATTCAGAGCGCGGAGACGAGTCCGAGTCTGCGCGCGCGCGCGTGAAGATGC 5039
 DB 4980 GGAATTCAGAGCGCGGAGACGAGTCCGAGTCTGCGCGCGCGCGCGTGAAGATGC 5039
 QY 5040 GCGCTCATGTATTCAGAGCGGCTTGTGCTTCACTGCGCGAGTCCGAGCTTGAAGCGCAT 5099
 DB 5040 GCGCTCATGTATTCAGAGCGGCTTGTGCTTCACTGCGCGAGTCCGAGCTTGAAGCGCAT 5099
 QY 5100 ACGCTTACCTGCGACTCTGTTCCGAGTGAAGATCAAGAGCGCATGAAGAACAA 5159
 DB 5100 ACGCTTACCTGCGACTCTGTTCCGAGTGAAGATCAAGAGCGCATGAAGAACAA 5159
 QY 5160 AATTAGTTTCTGTTGCTTGAAGAGTGTCCCTCCATGAGGTAAAGACTGTGATGA 5219
 DB 5160 AATTAGTTTCTGTTGCTTGAAGAGTGTCCCTCCATGAGGTAAAGACTGTGATGA 5219
 QY 5220 GTCTCAAGCTTACTGTTGAGTCTGCTGCGGTTGATTCATTCCCAAGAGCAAGAGG 5279
 DB 5220 GTCTCAAGCTTACTGTTGAGTCTGCTGCGGTTGATTCATTCCCAAGAGCAAGAGG 5279
 QY 5280 TGGCAACTAGTACGCGCGCGCTGAGATCA 5312
 DB 5280 TGGCAACTAGTACGCGCGCGCTGAGATCA 5312

RESULT 5
 AR125095 5368 bp DNA linear PAT 16-MAY-2001
 LOCUS AR125095
 DEFINITION Sequence 45 from patent US 6177075.
 ACCESSION AR125095
 VERSION AR125095.1 GI:14111157
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCES 1 (bases 1 to 5368)
 AUTHORS Christian P. Daniel., Gordon, K. Hienrich, Julius. and Hamlik, T. Nelson.
 TITLE Insect viruses and their uses in protecting plants
 JOURNAL Patent: US 6177075-A 45 23-JAN-2001;
 FEATURES
 source 1..5368
 BASE COUNT 1153 a 1747 c 1454 g 1014 t
 ORIGIN

Query Match 98.0%; Score 5206; DB 6; Length 5368;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 5312; Conservative 0; Mismatches 0; Indels 56; Gaps 5;
 QY 1 GTTCTGCTTCCCGGAGCGTAAATATAGGAGAACATGTACGCAAGAGCAGACGTTG 60
 DB 1 GTTCTGCTTCCCGGAGCGTAAATATAGGAGAACATGTACGCAAGAGCAGACGTTG 60
 QY 61 GGGCGGTGCTAGACCGCGCGGAGATGTCCTTACGCGGAGAGTCTGAGAGAGAGCAGT 120
 DB 61 GGGCGGTGCTAGACCGCGCGGAGATGTCCTTACGCGGAGAGTCTGAGAGAGAGCAGT 120
 QY 61 GGGCGGTGCTAGACCGCGCGGAGATGTCCTTACGCGGAGAGTCTGAGAGAGAGCAGT 120
 DB 61 GGGCGGTGCTAGACCGCGCGGAGATGTCCTTACGCGGAGAGTCTGAGAGAGAGCAGT 120
 QY 121 AAGTTGACTTGGCGCGCGCAGTGAAGGACATGAAAGCCCTCCAGAGACTGTACTATCG 180
 DB 121 AAGTTGACTTGGCGCGCGCAGTGAAGGACATGAAAGCCCTCCAGAGACTGTACTATCG 180
 QY 181 CTGCGCTTCAAGAGGAGGAGCTTTTACCCCGAGACAAACCGCATCTGCGCGGAGCA 240
 DB 181 CTGCGCTTCAAGAGGAGGAGCTTTTACCCCGAGACAAACCGCATCTGCGCGGAGCA 240

```

QY 241 CGTGTGACAGAGGTTCTGCACAATTTTCGCCAGGGGAGGTAGCAAGTGTGAGATA 300
Db 241 CGTGTGACAGAGGTTCTGCACAATTTTCGCCAGGGGAGGTAGCAAGTGTGAGATA 300
QY 301 GGGCCGCTCTTGACAGCGGCACTTAACTACATGGGGACCGAACCGCCCGTGTGAGAC 360
Db 301 GGGCCGCTCTTGACAGCGGCACTTAACTACATGGGGACCGAACCGCCCGTGTGAGAC 360
QY 361 TATCACGGGTGTACCAAGTACGGGACCGGCTGCGGACATTAACGCTTAAAG 420
Db 361 TATCACGGGTGTACCAAGTACGGGACCGGCTGCGGACATTAACGCTTAAAG 420
QY 421 TCTAGATTCGGTGCACAGGCGGGCCGAGTTCAAGGCGGACGCTACGCTGCGCAAC 480
Db 421 TCTAGATTCGGTGCACAGGCGGGCCGAGTTCAAGGCGGACGCTACGCTGCGCAAC 480
QY 481 GGGATTGCTCCCGGACCTTCTGCGTGTGACGAGTGGCTCTTGCGGTTCAATCGCG 540
Db 481 GGGATTGCTCCCGGACCTTCTGCGTGTGACGAGTGGCTCTTGCGGTTCAATCGCG 540
QY 541 GTTGAATTCGCAATCACTCCCTCTATGACGTGACCTTAAGAGAGCTGGCAATGCGTTT 600
Db 541 GTTGAATTCGCAATCACTCCCTCTATGACGTGACCTTAAGAGAGCTGGCAATGCGTTT 600
QY 601 GAGAACACGGAAGCTTCAATGATCCGCGGTTATGACATGACAGAGAGCTGTCTAC 660
Db 601 GAGAACACGGAAGCTTCAATGATCCGCGGTTATGACATGACAGAGAGCTGTCTAC 660
QY 661 ATGAGACACGTTGTTAATGCGGAGCTCGCTACGCTTCCAGTATTGAAAGAGCTATG 720
Db 661 ATGAGACACGTTGTTAATGCGGAGCTCGCTACGCTTCCAGTATTGAAAGAGCTATG 720
QY 721 GCTGTGAGAGACTGCGGATTCGAGGGGGGAGACTCGCTTCCAATTCCCTGAGTTGAG 780
Db 721 GCTGTGAGAGACTGCGGATTCGAGGGGGGAGACTCGCTTCCAATTCCCTGAGTTGAG 780
QY 781 TTTCATCAACGAGAGCCAAAGAGCGGCGATGAGAGAGCTGCGCGCGGCTCCTACTCC 840
Db 781 TTTCATCAACGAGAGCCAAAGAGCGGCGATGAGAGAGCTGCGCGCGGCTCCTACTCC 840
QY 841 AGACGCGCGCTATTTTCTCCGCGAGACGACGACTGGGTGATGACTTACAGACTTC 900
Db 841 AGACGCGCGCTATTTTCTCCGCGAGACGACGACTGGGTGATGACTTACAGACTTC 900
QY 901 CACACATGCTGCTCTTACTGATGAGGAACTAACCCCACTCGTTGGTTCTCACTC 960
Db 901 CACACATGCTGCTCTTACTGATGAGGAACTAACCCCACTCGTTGGTTCTCACTC 960
QY 961 CATATGAGAGTCCAGAGGGGCGCAAGGCTCCAGATTGAGTGGGACATTCGCGCGCA 1020
Db 961 CATATGAGAGTCCAGAGGGGCGCAAGGCTCCAGATTGAGTGGGACATTCGCGCGCA 1020
QY 1021 CCTGAGAGCGGAGTGTGCGCTGTCTCCAAAGAGAGCTCCAAAGGCTCTGCAAAATCCA 1080
Db 1021 CCTGAGAGCGGAGTGTGCGCTGTCTCCAAAGAGAGCTCCAAAGGCTCTGCAAAATCCA 1080
QY 1081 AACATCTTTTATTAACCGCGAGCGGTGGGACCTGAGAGATAAGACATCTTACGTACAG 1140
Db 1081 AACATCTTTTATTAACCGCGAGCGGTGGGACCTGAGAGATAAGACATCTTACGTACAG 1140
QY 1141 CACAAGTCAACATGCTGCTCAATTTTATGCAAAACGCTCTGAGAAAGAACTAGTCAG 1200
Db 1141 CACAAGTCAACATGCTGCTCAATTTTATGCAAAACGCTCTGAGAAAGAACTAGTCAG 1200
QY 1201 ATGACCGGTCTGATGTCGTTCCGCGCGGTAGGCTGGCGGATGCTGCTCAGAA 1260
Db 1201 ATGACCGGTCTGATGTCGTTCCGCGCGGTAGGCTGGCGGATGCTGCTCAGAA 1260
QY 1261 GTTACAGAGAGCTCTGAGAACTATCAACGGCTGACCTGCTCGCACTGTGCTCTT 1320
Db 1261 GTTACAGAGAGCTCTGAGAACTATCAACGGCTGACCTGCTCGCACTGTGCTCTT 1320
QY 1321 TAGCTCTCAATCATGAGCGCGGAGGGCTGCGGTGCTGCAAGACCGCAAGAG 1380
Db 1321 TAGCTCTCAATCATGAGCGCGGAGGGCTGCGGTGCTGCAAGACCGCAAGAG 1380

```

```

Db 1321 TAGCTCTCAATCATGAGCGCGGAGGGCTGCGGTGCTGCAAGACCGCAAGAG 1380
QY 1381 GACGTCTTTTGAAGAGCTTGTGTTCTGAGAGAGTCTCAAGCACTTGTGGCTCTGTG 1440
Db 1381 GACGTCTTTTGAAGAGCTTGTGTTCTGAGAGAGTCTCAAGCACTTGTGGCTCTGTG 1440
QY 1441 GGTCTGCGAAGCTTCAAGGACCGAGCTGTCTTTTACTTAAGCGCTGTGATTAAGTAC 1500
Db 1441 GGTCTGCGAAGCTTCAAGGACCGAGCTGTCTTTTACTTAAGCGCTGTGATTAAGTAC 1500
QY 1501 CGAGTCACTCGTGGAGAGCAATACTGAGAGTCCGCTGTCCCTGAAACAGTCCGC 1560
Db 1501 CGAGTCACTCGTGGAGAGCAATACTGAGAGTCCGCTGTCCCTGAAACAGTCCGC 1560
QY 1561 TTCCTGCGTCCCGCTACCACTGCGCGCTTCTTACAGACAGGAAAGTGTGAGTGC 1620
Db 1561 TTCCTGCGTCCCGCTACCACTGCGCGCTTCTTACAGACAGGAAAGTGTGAGTGC 1620
QY 1621 CTTCGGAAGCTGGCTGTCAACGAAAGTCCGCTTCACTCTCTCTGTGAGAGAG 1680
Db 1621 CTTCGGAAGCTGGCTGTCAACGAAAGTCCGCTTCACTCTCTCTGTGAGAGAG 1680
QY 1681 CCCCAAGTTTGAAGCGGAGCTTGTGAGAGAGCGGAGCGGCTCACTCCCGGAGTCCGC 1740
Db 1681 CCCCAAGTTTGAAGCGGAGCTTGTGAGAGAGCGGAGCGGCTCACTCCCGGAGTCCGC 1740
QY 1741 GCCACCTTGAGAGAGTCTCAACACCGAGCTCAAGACAGTCAAGTCACTCTGAGAA 1800
Db 1741 GCCACCTTGAGAGAGTCTCAACACCGAGCTCAAGACAGTCAAGTCACTCTGAGAA 1800
QY 1801 GGCCTTAAGACATGAGAGGCTCACTCTCTCCAGTCAAGAGGCTCGAGATGACAG 1860
Db 1801 GGCCTTAAGACATGAGAGGCTCACTCTCTCCAGTCAAGAGGCTCGAGATGACAG 1860
QY 1861 GGCCTCAAGACATGAGAGGCTCACTCTCTCCAGTCAAGAGGCTCGAGATGACAG 1920
Db 1861 GGCCTCAAGACATGAGAGGCTCACTCTCTCCAGTCAAGAGGCTCGAGATGACAG 1920
QY 1921 AAAGCACTTATGATGACACCAACAGAGAACTGAGAGAGCTATGAGACCGGAGTCAA 1980
Db 1921 AAAGCACTTATGATGACACCAACAGAGAACTGAGAGAGCTATGAGACCGGAGTCAA 1980
QY 1981 CGGCGTCCGCTCGGCTACGCAATGTGCGCTTGCATCTCCGTGTGCAACGCGC 2040
Db 1981 CGGCGTCCGCTCGGCTACGCAATGTGCGCTTGCATCTCCGTGTGCAACGCGC 2040
QY 2041 GAGGGGCGGCTTGTGCTACCGGTGTTATGAGAGTGTCAATGTTCCGCTCGGTAC 2100
Db 2041 GAGGGGCGGCTTGTGCTACCGGTGTTATGAGAGTGTCAATGTTCCGCTCGGTAC 2100
QY 2101 GTGCGATGTGACGCTGTGTCTCCGAGCTCAAGATAGTCTTGTAGGGAGGTCAC 2160
Db 2101 GTGCGATGTGACGCTGTGTCTCCGAGCTCAAGATAGTCTTGTAGGGAGGTCAC 2160
QY 2161 CAATGGGTTTATTAAGCTTCAAGGCAAGAGCGCAATGCGCTGTTGCGACGTC 2220
Db 2161 CAATGGGTTTATTAAGCTTCAAGGCAAGAGCGCAATGCGCTGTTGCGACGTC 2220
QY 2221 GTTAAAGCAGTGGCGTGGGCACTTTCAAGCAAGCAAGGCTGTGCGGCGAGCTGTT 2280
Db 2221 GTTAAAGCAGTGGCGTGGGCACTTTCAAGCAAGCAAGGCTGTGCGGCGAGCTGTT 2280
QY 2281 GCCACACGTTTTTCAAGAGCTTGTACCCCGGAGTGAACAACAACCTCAGGGTGTGCGCA 2340
Db 2281 GCCACACGTTTTTCAAGAGCTTGTACCCCGGAGTGAACAACAACCTCAGGGTGTGCGCA 2340
QY 2341 TCCATAGGCAAGTGGCGGCACTACCGGAAACAGGCGGCAAGCTCTGCTTCAAG 2400
Db 2341 TCCATAGGCAAGTGGCGGCACTACCGGAAACAGGCGGCAAGCTCTGCTTCAAG 2400
QY 2401 CAGAGAGAAAGTGGCGGCAAGGAGCTGAGGAGCGGATGACTGTGACAGAGCGGAG 2460

```

D 2401 CAGAGGAAAAAGTCGCCGCAACGGGGCTGAGGGCCGATGACTGTGCAAGAAACGACAGGA 2460
Q 2461 CGACCTTTGGTGTGTCAATTCTGCAATTACAAGGCTTCACAGACAGACAGAACTCTC 2520
D 2461 CGCACTTTTGGCTGTGCAATTCTGCAATTACAAGGCTTCACAGACAGACAGAACTCTC 2520
Q 2521 GCTGAGAAAGTCGACCTTCTAGTGGGATCAACGGCCACACCAACCACTTATATCCG 2580
D 2521 GCTGAGAAAGTCGACCTTCTAGTGGGATCAACGGCCACACCAACCACTTATATCCG 2580
Q 2581 GACCCGACAGGTGACATTGAGACACTCAACCTAGCGGAAAGCGAGGTGTTTACA 2640
D 2581 GACCCGACAGGTGACATTGAGACACTCAACCTAGCGGAAAGCGAGGTGTTTACA 2640
Q 2641 GACATCCCTGCAACCCCTGAGATCAACGCTGCAACCGAGTGAAGAGTGCAGCGCAAC 2700
D 2641 GACATCCCTGCAACCCCTGAGATCAACGCTGCAACCGAGTGAAGAGTGCAGCGCAAC 2700
Q 2701 GAATGATGGAAGATACCCCGGAGATGCAACCGCGGACAGAGCAATCCATCTGCTC 2760
D 2701 GAATGATGGAAGATACCCCGGAGATGCAACCGCGGACAGAGCAATCCATCTGCTC 2760
Q 2761 CGCAAGAACTTCGGGGACCAACCGGACTGTGTGCTTGGCGAAGACCGGCTAC 2820
D 2761 CGCAAGAACTTCGGGGACCAACCGGACTGTGTGCTTGGCGAAGACCGGCTAC 2820
Q 2821 GAGGTGTTTGGCGGTGCTGCAAAAATCAACGTAAGCTTTCGAAACCGGACCGAC 2880
D 2821 GAGGTGTTTGGCGGTGCTGCAAAAATCAACGTAAGCTTTCGAAACCGGACCGAC 2880
Q 2881 AAGCCGATAGGGGCTTCAGAGAGGGGTACAGTGGTCAAGGTCAACGCGCTTAC 2940
D 2881 AAGCCGATAGGGGCTTCAGAGAGGGGTACAGTGGTCAAGGTCAACGCGCTTAC 2940
Q 2941 AAACACGAGCGCTCCAGACGCTGTTGCCGCTACACCAAGCGAAGGCTGACCTGCG 3000
D 2941 AAACACGAGCGCTCCAGACGCTGTTGCCGCTACACCAAGCGAAGGCTGACCTGCG 3000
Q 3001 CTACACGAAAGCTAAGAGAGAGCTCAACGCAATGCTAACTCGCTTGAACGACATTGGAG 3060
D 3001 CTACACGAAAGCTAAGAGAGAGCTCAACGCAATGCTAACTCGCTTGAACGACATTGGAG 3060
Q 3061 TGGAGTGTCACTGAAGAGCCCGGTGACCGAGCTGTCTGAGACCCAGCTCAAGTTACC 3120
D 3061 TGGAGTGTCACTGAAGAGCCCGGTGACCGAGCTGTCTGAGACCCAGCTCAAGTTACC 3120
Q 3121 CAACGCGCGGACCGCTGCAAGAGCTGTGAGAGCAGACCCCTTACATCGGTGACATA 3180
D 3121 CAACGCGCGGACCGCTGCAAGAGCTGTGAGAGCAGACCCCTTACATCGGTGACATA 3180
Q 3181 GACTTCCTTATGAAGACTCAGACAGAAAGTGTCCCAAGCGATCAATACGGGCAAGTTC 3240
D 3181 GACTTCCTTATGAAGACTCAGACAGAAAGTGTCCCAAGCGATCAATACGGGCAAGTTC 3240
Q 3241 GGGGAGGGGATTCGCGCTCACTCAAAAGTCTCAACTTCGCTGCGCTTGGATAGC 3300
D 3241 GGGGAGGGGATTCGCGCTCACTCAAAAGTCTCAACTTCGCTGCGCTTGGATAGC 3300
Q 3301 ATAATCGAGAGATCTCGTACCGGAGCGGACAGTCCGGTACAGCAAGGTTCGCC 3360
D 3301 ATAATCGAGAGATCTCGTACCGGAGCGGACAGTCCGGTACAGCAAGGTTCGCC 3360
Q 3361 GACGAGAAAGAGGCGCATCTGCTGAGAGCAAGATCAATCAAGTCCCAACGCGCTTC 3420
D 3361 GACGAGAAAGAGGCGCATCTGCTGAGAGCAAGATCAATCAAGTCCCAACGCGCTTC 3420
Q 3421 GTCTCGGGGACTGAGACCGAGTTTGAACCGGCCCAATTAACAGAGTGTCTTTC 3480
D 3421 GTCTCGGGGACTGAGACCGAGTTTGAACCGGCCCAATTAACAGAGTGTCTTTC 3480
Q 3481 GCCGCCCTTTTGAAGCGCATCGGACCGCTGCAAGTCTGCTTATCTTATCAGAGAACGG 3540
D 3481 GCCGCCCTTTTGAAGCGCATCGGACCGCTGCAAGTCTGCTTATCTTATCAGAGAACGG 3540

Q 3541 TGTGGAAAAAGCACTTTCGAGAGGAGGTCTAGGCTTCGTTGAAGTCAACGCTCTGCTC 3600
D 3541 TGTGGAAAAAGCACTTTCGAGAGGAGGTCTAGGCTTCGTTGAAGTCAACGCTCTGCTC 3600
Q 3601 GACTCCGGGAGCTTGAAGCGCCCTTGGCGCAACCACTTCTCTGCGCGGTCAATGCTC 3660
D 3601 GACTCCGGGAGCTTGAAGCGCCCTTGGCGCAACCACTTCTCTGCGCGGTCAATGCTC 3660
Q 3661 ACGCTCTTCGCGCGGCTCAAGTTTCAGAGCTTTCAGAGCGACCTGCTCTGTGTGT 3720
D 3661 ACGCTCTTCGCGCGGCTCAAGTTTCAGAGCTTTCAGAGCGACCTGCTCTGTGTGT 3720
Q 3721 AGCCATTACTCCGTTTCAGACGCTAGCCGCTTCACTGGGCGGAAAGTTTACAGACAAA 3780
D 3721 AGCCATTACTCCGTTTCAGACGCTAGCCGCTTCACTGGGCGGAAAGTTTACAGACAAA 3780
Q 3781 CATTGAAAGGTGAGAGTGAAGAAAAATGTCACCGTACATCGGACTCTCGTCCGCTGAG 3840
D 3781 CATTGAAAGGTGAGAGTGAAGAAAAATGTCACCGTACATCGGACTCTCGTCCGCTGAG 3840
Q 3841 CAAGTCGTCTGCAACCCCTGTCAAGAGCGCTCTCAAGATATTGGGCGCTGCTACAAAC 3900
D 3841 CAAGTCGTCTGCAACCCCTGTCAAGAGCGCTCTCAAGATATTGGGCGCTGCTACAAAC 3900
Q 3901 GAATCTCTTTTACTCAAGTACGTGAGAGCTGTGAGAGACATCACCAAGGCTGAGATGAC 3960
D 3901 GAATCTCTTTTACTCAAGTACGTGAGAGCTGTGAGAGACATCACCAAGGCTGAGATGAC 3960
Q 3961 GCCCGCTACCAAGCCCTCTGTGCAATGTCAGAGCATGCTTACATTAACGCGCGGAG 4020
D 3961 GCCCGCTACCAAGCCCTCTGTGCAATGTCAGAGCATGCTTACATTAACGCGCGGAG 4020
Q 4021 TCTGCGCGGTACATCATGACGCTGTGTTGCTGCTTGGGCGCGGCGACTTCCGTTTGA 4080
D 4021 TCTGCGCGGTACATCATGACGCTGTGTTGCTGCTTGGGCGCGGCGACTTCCGTTTGA 4080
Q 4081 CAATCGCGCGGTGTGCTGCTCATGTGCAAGACCCGACCGCTTACAGACAGACGTATCCG 4140
D 4081 CAATCGCGCGGTGTGCTGCTCATGTGCAAGACCCGACCGCTTACAGACAGACGTATCCG 4140
Q 4141 GCTAACGTCGCGCATGTCGCTTGAACAGCTCTTGAAGCCCGGACGCGCGCGCCG 4200
D 4141 GCTAACGTCGCGCATGTCGCTTGAACAGCTCTTGAAGCCCGGACGCGCGCGCCG 4200
Q 4201 GCAAGTTTCGTTGCAATGTGCAACCGGAAACGCTTCTTCACTTACCGCGAAAGCT 4260
D 4201 GCAAGTTTCGTTGCAATGTGCAACCGGAAACGCTTCTTCACTTACCGCGAAAGCT 4260
Q 4261 GGTGTTTCTGCACTACAGGCACTGTGCGACTGCGGCGCGGAGTCTTCATGG 4320
D 4261 GGTGTTTCTGCACTACAGGCACTGTGCGACTGCGGCGCGGAGTCTTCATGG 4320
Q 4321 GATGACCTGAGGCAACAGTTTTCGGAAGTTTGAACCGGAGACCCCGTCCACATCA 4380
D 4321 GATGACCTGAGGCAACAGTTTTCGGAAGTTTGAACCGGAGACCCCGTCCACATCA 4380
Q 4381 TCTTCGCGGTATCGTCTTTCAGACTCGACTCTGATGTAAGTTCGCTCATGTGT 4440
D 4381 TCTTCGCGGTATCGTCTTTCAGACTCGACTCTGATGTAAGTTCGCTCATGTGT 4440
Q 4441 GGAAGACCCGAGAGACACAGAAAGCTTGAACAGGCAAGGCGCTTTCGACAGACAG 4500
D 4441 GGAAGACCCGAGAGACACAGAAAGCTTGAACAGGCAAGGCGCTTTCGACAGACAG 4500
Q 4501 CAATCAGCGCTGTGTAATGTGAGACAGAAAGCGGAAAGGACAGGAGATTTCGTTACT 4560
D 4501 CAATCAGCGCTGTGTAATGTGAGACAGAAAGCGGAAAGGACAGGAGATTTCGTTACT 4560
Q 4561 GCCCCACTGTCTCGAGCCCTCATTTCTCAATTTTCGAAAGAGCTGACATGGACACCGG 4620
D 4561 GCCCCACTGTCTCGAGCCCTCATTTCTCAATTTTCGAAAGAGCTGACATGGACACCGG 4620

QY 4621 CGCAGTGTGCGCGCTGAGATCATCTTGGGCAACCCCATCTGCGGCAAGCAAGTT 4680
 DB 4621 CCACAGTGTGCGCGCTGAGATCATCTTGGGCAACCCCATCTGCGGCAAGCAAGTT 4680
 QY 4681 GCGCGAGAGACACAGCGCGGAGCTTTCCTTCTGAGGATCCCAAGTGTGCGCGGATGC 4740
 DB 4681 GCGCGAGAGACACAGCGCGGAGCTTTCCTTCTGAGGATCCCAAGTGTGCGCGGATGC 4740
 QY 4741 TCGAAGCGGATCCGCGCGCGCGCGAGCTGCGGTTGGAAGAAGTCAACCGCGCTCCACGCG 4800
 DB 4741 TCGAAGCGGATCCGCGCGCGCGCGAGCTGCGGTTGGAAGAAGTCAACCGCGCTCCACGCG 4800
 QY 4801 TGAAGGCGGTGACCGCGAGACCGAGAGAGTCAAGGAGAGACCGGAGACAGCGCGGTG 4860
 DB 4801 TGAAGGCGGTGACCGCGAGACCGAGAGAGTCAAGGAGAGACCGGAGACAGCGCGGTG 4860
 QY 4861 GTCGAGGCTGTGATCAGCGCGCGCTTATCTCAGAGAAGAGCTTTCCTCCAGCGATCC 4920
 DB 4861 GTCGAGGCTGTGATCAGCGCGCGCTTATCTCAGAGAAGAGCTTTCCTCCAGCGATCC 4920
 QY 4921 AAGGCTACTCAAGAACTAAGAGAGTCTCAACATCCACTCTTCTCTG---CCCGAGT 4977
 DB 4921 AAGGCTACTCAAGAACTAAGAGAGTCTCAACATCCACTCTTCTCTG---CCCGAGT 4980
 QY 4978 GCGGATTTACAGGCGCGCGAGCTGCGAGACAGTCCGAGTCTG-----CCGCGCGCTGC 5030
 DB 4981 GCGGATTTACAGGCGCGCGAGCTGCGAGACAGTCCGAGTCTGCGGAGACCGCGCGCTGC 5040
 QY 5031 AGAGATGCGCGCGCTCATGTATTCACGAGCGGTTGGC-----TTTCAT 5071
 DB 5041 AGAGATGCGCGCGCTCATGTATTCACGAGCGGTTGGC-----TTTCAT 5100
 QY 5072 CTGCGCGCGAGTGCAGCTTGAAGCGCATGCTCTTACCTCGA-----CTCT 5118
 DB 5101 CTGCGCGCGAGTGCAGCTTGAAGCGCATGCTCTTACCTCGA-----CTCT 5160
 QY 5119 GTTCCCGATTAAGATCAGCAAGAGCGCATGGAAGAAACAAAT-----TA 5164
 DB 5161 GTTCCCGATTAAGATCAGCAAGAGCGCATGGAAGAAACAAAT-----TA 5220
 QY 5165 GTTCCCTGTTGTTAAACAAGAGTGTCTCCCTCCCATTTAGGTTAAAGATCTGTGAGTCT 5224
 DB 5221 GTTCCCTGTTGTTAAACAAGAGTGTCTCCCTCCCATTTAGGTTAAAGATCTGTGAGTCT 5280
 QY 5225 CAACGTTACTCGTTGAGTGTCTGCGGTTGATTCATTCCCAAGCAGCAAGAGGTGCGC 5284
 DB 5281 CAACGTTACTCGTTGAGTGTCTGCGGTTGATTCATTCCCAAGCAGCAAGAGGTGCGC 5340
 QY 5285 AACTAGTACGCGCGCGCGCGCGCTGGATACCA 5312
 DB 5341 AACTAGTACGCGCGCGCGCGCGCTGGATACCA 5368

RESULT 6
 AF102884 6625 bp RNA linear VRL 30-MAY-2001
 LOCUS Nudaurelia capensis beta virus complete genome.
 DEFINITION AF102884
 ACCESSION AF102884.1 GI:4028586
 VERSION
 KEYWORDS Nudaurelia capensis beta virus
 SOURCE Nudaurelia capensis beta virus
 ORGANISM Betatetravirus.
 1 (bases 1 to 6625)
 Gordon, K.H., Williams, M.R., Hendry, D.A. and Hanzlik, T.N.
 TITLE Sequence of the genomic RNA of nudaurelia beta virus (Tetraviridae)
 JOURNAL Virology 258 (1), 42-53 (1999)
 MEDLINE 99263183
 PUBMED 10329566
 REFERENCE 2 (bases 1 to 6625)
 Gordon, K.H., Williams, M.R., Hendry, D.A. and Hanzlik, T.N.

TITLE Direct Submission
 JOURNAL Submitted (30-OCT-1998) CSIRO Entomology, GPO Box 1700, Canberra, ACT 2611, Australia
 FEATURES
 source
 CDS
 1..6625
 /organism="Nudaurelia capensis beta virus"
 /mol_type="genomic RNA"
 /db_xref="taxon:85652"
 93..5870
 /note="replicase"
 /codon_start=1
 /product="RNA-dependent RNA polymerase"
 /protein_id="AAC97509.1"
 /db_xref="GI:4028587"
 /translation="MEDAKKQRLVLDQERAKAFLQDLDFASVETLEDAQEKYGMF
 RSGTKLPSTHIXLALDYAEKDLRRHVKNVPLVLEIGSVESVRAVQDERYMG
 CFSFDARNDLRNKKIGYEAYRCKIPDAALAAQIPTPFCVDGSGNCEYQSLIA
 CHSLYVDGSGNSIMDYAKGMALHGTIVYAMMLPEELLITPDADNIPFGYSIRPET
 GALPCTRRKKAIPSGYNDGSAVYDHANHACGLKRGCDTPGFSILIDIQRFGM
 TKLITRHSQSIITVPLSKGLIWNENIYKIMPKAKHEBYIVTDKKEGYCV
 YGTRVQSGKSIITLAEIYOYIRTLRLIINGTVAHKWTALADIERLAVISIMFK
 NERAVSERALMRAQCKKSAEKQALLPYMERIANMPQDKQIDEEVARKLECLK
 AOPWIMADRVNCEKRYNPYAEVGRKHLATGLRELQREIPANEPQDGAAYM
 HSNADLDIYAEGLIDSAKEAAKQSLAITLQQAFOYLGRKTCGCGNNIEBYTG
 PPSGSGRAKRPADLQGVLCAPTRLRLDLDSVHPSTKCTYHNALHYAAKES
 GNRPEVIVIDEKTPACVGTMHASPSRSLVCGDHPQIGYIDPSRKDLKPS
 IIAECRTKRPFTYTRCPQDVNLPIFKTLYPALSPSKQLSIRLTALVATYRTH
 AQLTQDDKPHSPVPTAHPQARLTVHAGTLPBRLALEKHIALVIRHRTNA
 LYIRDSSEKGLVPLSLMTPPMSYTCVTDQWQVLDGVPVAVRENSSGPDSHNGA
 ITLQELGLTPTKRYRPESEAVPTAHNRVLTAKLDELGDPBRYEPTQFNLRGCTY
 NIKDQALHTLVGRYARKINSRREDARFVAKITAKLKEWIPPTAEBOVDSCA
 DAMQKIAERGHVDIEDPWSNEGRIIYHLKQOQVMPDTKIKLQGISAKHCANI
 ALASAVRIITODMSSEKFIANGQSDRPTDSIIEARLOKAEKFSIDIKERDTYN
 WSLIVFSRRCDCGCEHLIEYFEKSKRTLSRSIGVDSFMLDSGAWMTIARNTL
 PASGMLALFVGQDPIAAKGDDVFLAGNNLIDAEKRLMGSYLAANNIKETAAVSE
 IGVISQAAVTAADVPLATRTYGRSYKNNDLAKYIAIADHCKIPRSRPTMTAIN
 CATLYTSKECTINYLDALDARGHTMSDLIDPGVMTYTPKVERVTSQDQGR
 ADKTRERKQPEQPPQOQOASTOAGSKTSRSTVDYOPRDGTRERERPGDQR
 SDTRBQVKAISDDESHSDIRGDSRLSPGRITIOBEPRENSRRDTSVNNRSISR
 DGRQIPGTEPYDALTAGMRDLASTDSASVPLQASVVVHHHQHQSDEBRGCVRE
 RLEQOQGLRSQVPLKAGSERVYLRHGRPRSDAGRTTPSTGCTIRTRLEPISIBRRH
 SVLQRTHSRSGSGDRAVPTGQRTPEGRSGSRPHGRNGYTAARFAELIIDD
 RQGRVYRGSNPATHTGVADADAGVGAAGTCDPCEPDIRRKNHNDHAAATRVGDVYA
 IHSQQRDHRDRDGSATVRSSEFLRLTESAGHQLNDSTNEHSPNAGNAQDSHV
 PTOENGLIYAEQGVPTRVANDNGDPIYVAGMEIDTDCGRPLPGMMWTFGYRQDLR
 DRCRDQRYVYINRLTLOGVPTLRSDPTGSGEPGLQPCDTSGRGANSGLDLSA
 PIRPCTGIRIGPIRDGGDHSPTSLCAISRSBCECDGLHREDRDRECSLEPHLGE
 AATKSETCRNRSRPSQCGPHREP"
 4039..5877
 /note="coat protein; forms T=4 capsids with 240 copies of
 protein per particle"
 /codon_start=1
 /product="capsid protein"
 /protein_id="AAC97510.1"
 /db_xref="GI:4028588"
 /translation="MDANVOIRPARNPQSGNRRNNKRRRRRLGLKLPVPAVPT
 SPQMAEPANHANTRVNRGRTRGLRQMAEMSPMAATSBAMHDVLDGSEKTSID
 DGRIPDGAIPSTCGQFRTGVGARYPGLNSTLPLDGGTWPMLVLMPLPFRHPLFTT
 TTSTNEVEVTNADLIDAPANDMNRNTDWTATYPSMAQVNGVPMVYPTALITLPPPT
 QLQVSGLLSYRLTSGYATYRAAPTLVAVQAVVIAQFQPDREKQENPDIVAGTQY
 GGLQIQSGSPNTLTMTIGDOVERFGAALPLPTVSGMPSPSGQLVPOFANLTFSG
 NNTITITLTPSGSVTGMQFTASNGTDTVYVAGALVAFGALVLDSEINLQDINSG
 IPEPTNQMOMQADTPTKIPOLEMETKGFVPLPQCVFETMTVATSGVPMVKPLRTV
 VDHRAIIGLOLQADTIDENPAIGVAMTKGMSLTPVYKVRFRPAIDAGSGPMPASA
 TPQKQVALTVARWTDLHPAYPERYVNGALFAMVARTIAQIPYVSAAGVANA
 TQICBSATRESVANSNTERRRARRARVGIARARLVGRIGLST"
 5644..5649
 /note="cleavage site for coat protein after assembly"
 6514..6625
 /note="tRNA-like structure; valine anticodon; no
 pseudo-knot in the aminoacyl stem"

BASE COUNT 1696 a 1800 c 1783 g 1346 t
 ORIGIN

TITLE Modified small RNA viruses
JOURNAL Patent: US 6251654-A 4 26-JUN-2001;
FEATURES Location/Qualifiers
SOURCE 1. .6534
/organism="unknown"

BASE COUNT 1667 a 1782 c 1747 g 1338 t

ORIGIN

Query Match 1.44; Score 76.4; DB 6; Length 6534;
Best Local Similarity 52.5%; Pred. No. 0.00026;
Matches 192; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 3546 GAAACGACCTTGGAGCGAAGGCTAGCGCTCGGTAAGTGAACGAGTCTGCGATC 3605
DB 3367 GAAAGCGGACGCTCTCAAGCCCATAGAGAGCGTCACTTACCTTACGCTCGATTC 3426
QY 3606 CGGCGACGCTTGAAGCGCTTGCAGCAACCATCTTCTGTCGCGCGATGCTCAGCT 3665
DB 3427 TGGCGCTGTGAGCAATTCGCAAGAACCTTATTTGCTCGGGTCTTATGCTCGCCCT 3486
QY 3666 CTTCGCGGCGCTCAAGTTGCGAGCTTTCAAGGCGAGCACTGCTCTGTGTAGCCA 3725
DB 3487 TTTCGTGCGGCTCATTTTCATCGCGGCAAGGCGATGATGTTCTCTCGAGGGA--A 3543
QY 3726 TTACCTCGCTTGGACCGCTTACATGAGGGAAGCTTACAAGACCAACATTT 3785
DB 3544 TAACTTGTACTTGGACGAGAACGCGCTTCGATGAGATCTTACCTTACCGCAACACTT 3603
QY 3786 GAAAGTCGAGTGAAGAAATCGTCGCTACATCGACTCCTCGCTCCGCTGAGCAGT 3845
DB 3604 GAAGATGAGAAAGACGCGGCTGAGAGCTTATAGGTTATCGTTTCCCAAGCGCGCT 3663
QY 3846 CGTCCTGACCTCTCAAGAGCGCTCTCAAGATATTGGGCGCTGCTACACAGCGAAT 3905
DB 3664 CACAGCTGATGTCTGCGTACGACCGGACTTACGCTGAGATATAAAGCGTAT 3723
QY 3906 CCTTTA 3911
DB 3724 GATCTA 3729

RESULT 8
AX028953
LOCUS AX028953 6536 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 4 from Patent WO9746666.
ACCESSION AX028953
VERSION AX028953.1 GI:10189994
KEYWORDS
SOURCE Nudaurelia capensis beta virus
ORGANISM Nudaurelia capensis beta virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae; Betatelevirus.

REFERENCE
1 Gordon, K.H. and Hanzlik, T.N.
TITLE Modified small rna viruses
JOURNAL Patent: WO 9746666-A 4 11-DEC-1997;
COMM SCIENT IND RES ORG (AU); GORDON KARL HEINRICH (AU); HANZLIK TERRY NELSON (AU)
FEATURES Location/Qualifiers
source 1. .6536
/organism="Nudaurelia capensis beta virus"
/mol_type="genomic DNA"
/db_xref="taxon:85652"

BASE COUNT 1667 a 1782 c 1747 g 1338 t 2 others

ORIGIN

Query Match 1.44; Score 76.4; DB 6; Length 6536;
Best Local Similarity 52.5%; Pred. No. 0.00026;
Matches 192; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 3546 GAAACGACCTTGGAGCGAAGGCTTAGCGCTCGGTAAGTGAACGAGTCTGCGATC 3605
DB 3369 GAAAGCGGACGCTCTCAAGCCCATAGAGAGCGTCACTTACCTTACGCTCGATTC 3428

QY 3606 CGGCGACGCTTGAAGCGCTTGCAGCAACCATCTTCTGCGCGCTCATGCTCAGCT 3665
DB 3429 TGGCGCTGTGAGCCATTCACGAAACACCTTATTTGCTCGGGTCTTATGCTCGCCCT 3488
QY 3666 CTTCGCGGCGCTCAAGTTTCCAGCTTCAAGGCGAGCACTCGCTCGCTGTGTAGCCA 3725
DB 3489 TTTCGTGCGGCTGATTTTCATTCGCGGCAAGGCGATGATGTCCTTCTCGAGGGA--A 3545
QY 3726 TTACCTCGCTTGCAGCGCTAGCGCGCTTCAATAGGCGAAAGTTACAAAGCAACATTT 3785
DB 3546 TAACTTGTACTTGGACGCAAGCGCTTCGATGAGATCTTACTTACCGCAACACTT 3605
QY 3786 GAAAGTCGAGTGAAGAAATCGTCGCTACATGCACTCTGCTCTCGCTGAGCAGT 3845
DB 3606 GAAGATGAGAAAGACGCGGCTGAGACTTATAGGGTTATAGCTTTTCCCAAGCGCGCT 3665
QY 3846 CGTCCTGACCGCTGTGAGAGCGCTCTCAAGATATTGGGCGCTGCTACAAAGCAACT 3905
DB 3666 CACAGCTGATGTCTGCGTACGACCGGACTTACGCTGAAAGTTATATAAGCGTAT 3725
QY 3906 CCTTTA 3911
DB 3726 GATCTA 3731

RESULT 9
AX028972/c
AX028972 6536 bp DNA linear PAT 16-SEP-2000
DEFINITION Sequence 23 from Patent WO9746666.
ACCESSION AX028972
VERSION AX028972.1 GI:10190010
KEYWORDS
SOURCE Nudaurelia capensis beta virus
ORGANISM Nudaurelia capensis beta virus
Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae; Betatelevirus.

REFERENCE
1 Gordon, K.H. and Hanzlik, T.N.
TITLE Modified small rna viruses
JOURNAL Patent: WO 9746666-A 23 11-DEC-1997;
COMM SCIENT IND RES ORG (AU); GORDON KARL HEINRICH (AU); HANZLIK TERRY NELSON (AU)
FEATURES Location/Qualifiers
source 1. .6536
/organism="Nudaurelia capensis beta virus"
/mol_type="genomic DNA"
/db_xref="taxon:85652"

BASE COUNT 1338 a 1747 c 1782 g 1667 t 2 others

ORIGIN

Query Match 1.44; Score 76.4; DB 6; Length 6536;
Best Local Similarity 52.5%; Pred. No. 0.00026;
Matches 192; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 3546 GAAACGACCTTGGAGCGAAGGCTTAGCGCTCGGTAAGTGAACGAGTCTGCGATC 3605
DB 3168 GAAAGCGGACGCTCTCAAGCCCATAGAGAGCGTGAAGTGAAGTCACTCGATTC 3109
QY 3606 CGGCGACGCTTGAAGCGCTTGCAGCAACCATCTTCTGCGCGCTCATGCTCAGCT 3665
DB 3108 TGGCGCTGTGAGCAATTCGCAAGAACCTTATTTGCTCGGGTCTTATGCTCGCCCT 3049
QY 3666 CTTCGCGGCGCTCAAGTTTCCAGCTTCAAGGCGAGCACTGCTCTGTGTAGCCA 3725
DB 3048 TTTCGTGCGGCTGATTTTCATTCGCGGCAAGGAGATGATGTCCTCTCGAGGGA--A 2992
QY 3726 TTACCTCGCTTGAAGCTTACCGCGCTTCAATGCGGAAAGTTTACAAGACCAACATTT 3785
DB 2991 TAACTTGTACTTGGACGCAAGCGCTTCAATGAGATCTTACTTACCGCAACACTT 2932
QY 3786 GAAAGTCGAGTGAAGAAATCGTCGCTACATGCACTCTGCTCGCTGAGCAGT 3845

DB 2931 GAAGATCGAAGAAAGCGGGCTGTAGCTTTATAGGTTTATCGTTTCCCAAGCCGCCGT 2872

QY 3846 CGTCCCTGACCCCTGTACAGAGCGCTCTCAAGATATTGGCGGCTCTACACAGCGAACT 3905

DB 2871 CACAGCTGATGTGCGCTGTAGCCACCGGACTTACGATCGAAGTTATTAAGAAAGTGAT 2812

QY 3906 CCTTA 3911

DB 2811 GATCTA 2806

RESULT 10

ARI25096

LOCUS ARI25096 2478 bp DNA linear PAT 16-MAY-2001

DEFINITION Sequence 47 from patent US 6177075.

ACCESSION ARI25096

VERSION ARI25096.1 GI:14111158

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2478)

AUTHORS Christian,P.Daniel., Gordon,K.Hienrich.Julius. and Hanzlik,T.Nelson.

TITLE Insect viruses and their uses in protecting plants

JOURNAL Patent: US 6177075-A 47 23-JAN-2001;

FEATURES

source Location/Qualifiers

1..2478

/organism="unknown"

BASE COUNT 516 a 854 c 607 g 501 t

ORIGIN

Query Match 1.3%; Score 71.2; DB 6; Length 2478;

Best Local Similarity 78.7%; Pred.No. 0.0028;

Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 5205 TAAAGACTGTGTAGTCTCTCAAGTACTGCTGAGTGTGCGGTTGCATTCCATTTC 5264

DB 2371 TAAAGACTGTGTAGTCTCTCTCAAGTACTGCTGAGTGTGCGGTTGCATTCCATTTC 2430

QY 5265 CCAAGCAGCAAGGAGGTCGCACTAGTACGGGCCCTGGGATACCA 5312

DB 2431 CCAAGCGGCAAGAGAGAGCTAGTACTGCTGCTGCCCTCGGATACCA 2478

RESULT 11

ARI25097

LOCUS ARI25097 2478 bp DNA linear PAT 16-MAY-2001

DEFINITION Sequence 49 from patent US 6177075.

ACCESSION ARI25097

VERSION ARI25097.1 GI:14111159

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2478)

AUTHORS Christian,P.Daniel., Gordon,K.Hienrich.Julius. and Hanzlik,T.Nelson.

TITLE Insect viruses and their uses in protecting plants

JOURNAL Patent: US 6177075-A 49 23-JAN-2001;

FEATURES

source Location/Qualifiers

1..2478

/organism="unknown"

BASE COUNT 516 a 854 c 607 g 501 t

ORIGIN

Query Match 1.3%; Score 71.2; DB 6; Length 2478;

Best Local Similarity 78.7%; Pred.No. 0.0028;

Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 5205 TAAAGACTGTGTAGTCTCTCAAGTACTGCTGAGTGTGCGGTTGCATTCCATTTC 5264

DB 2371 TAAAGACTGTGTAGTCTCTCTCAAGTACTGCTGAGTGTGCGGTTGCATTCCATTTC 2430

QY 5265 CCAAGCAGCAAGGAGGTCGCACTAGTACGGGCCCTCGGATACCA 5312

DB 2431 CCAAGCGGCAAGAGAGAGTGTAGTACTGCTGCTGCCCTCGGATACCA 2478

RESULT 12

ARI59968

LOCUS ARI59968 2478 bp DNA linear PAT 17-OCT-2001

DEFINITION Sequence 1 from patent US 6251654.

ACCESSION ARI59968

VERSION ARI59968.1 GI:16222852

KEYWORDS

SOURCE Unknown.

ORGANISM Unknown.

REFERENCE 1 (bases 1 to 2478)

AUTHORS Gordon,K.Hienrich. and Hanzlik,T.Nelson.

TITLE Modified small RNA viruses

JOURNAL Patent: US 6251654-A 1 26-JUN-2001;

FEATURES

source Location/Qualifiers

1..2478

/organism="unknown"

BASE COUNT 516 a 854 c 607 g 501 t

ORIGIN

Query Match 1.3%; Score 71.2; DB 6; Length 2478;

Best Local Similarity 78.7%; Pred.No. 0.0028;

Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 5205 TAAAGACTGTGTAGTCTCTCAAGTACTGCTGAGTGTGCGGTTGCATTCCATTTC 5264

DB 2371 TAAAGACTGTGTAGTCTCTCTCAAGTACTGCTGAGTGTGCGGTTGCATTCCATTTC 2430

QY 5265 CCAAGCAGCAAGGAGGTCGCACTAGTACGGGCCCTCGGATACCA 5312

DB 2431 CCAAGCGGCAAGAGAGAGTGTAGTACTGCTGCTGCCCTCGGATACCA 2478

RESULT 13

AX028950

LOCUS AX028950 2478 bp DNA linear PAT 16-SEP-2000

DEFINITION Sequence 1 from Patent WO9746666.

ACCESSION AX028950

VERSION AX028950.1 GI:10189993

KEYWORDS

SOURCE Helicoverpa armigera stunt virus

ORGANISM Helicoverpa armigera stunt virus

REFERENCE 1

AUTHORS Gordon,K.H. and Hanzlik,T.N.

TITLE Modified small rna viruses

JOURNAL Patent: WO 9746666-A 1 11-DEC-1997;

COMMUN SCIENT IND RES ORG (AU) ; GORDON KARL HEINRICH (AU) ; HANZLIK TERRY NELSON (AU)

FEATURES

source Location/Qualifiers

1..2478

/organism="Helicoverpa armigera stunt virus"

/mol_type="genomic DNA"

/db_xref="taxon:37206"

BASE COUNT 516 a 854 c 607 g 501 t

ORIGIN

Query Match 1.3%; Score 71.2; DB 6; Length 2478;

Best Local Similarity 78.7%; Pred.No. 0.0028;

Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 5205 TAAAGACTGTGTAGTCTCTCAAGTACTGCTGAGTGTGCGGTTGCATTCCATTTC 5264

DB 2371 TAAAGACTGTGTAGTCTCTCTCAAGTACTGCTGAGTGTGCGGTTGCATTCCATTTC 2430

QY 5265 CCAAGCAGCAAGGAGGTCGCACTAGTACGGGCCCTCGGATACCA 5312


```

Db      2431 CCAAGCGGCAAGAGCAGTAGTACTGCTGCGCTCGCGGATACCA 2478
|||||
RESULT 14
LOCUS   AX028971/c      2478 bp      DNA      linear      PAT 16-SEP-2000
DEFINITION Sequence 22 from Patent WO9746666.
ACCESSION AX028971
VERSION   AX028971.1 GI:10190009
KEYWORDS
SOURCE   Helicoverpa armigera stunt virus
ORGANISM Helicoverpa armigera stunt virus
          Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
          unclassified Tetraviridae.
REFERENCE
AUTHORS Gordon,K.H. and Hanzlik,T.N.
TITLE     Modified small rna viruses
JOURNAL   Patent: WO 9746666-A 22 11-DEC-1997;
          COMMW SCIENT IND RES ORG (AU) ; GORDON KARL HEINRICH (AU) ; HANZLIK
          TERRY NELSON (AU)
FEATURES
         source
         1..2478
            /organism="Helicoverpa armigera stunt virus"
            /mol_type="genomic DNA"
            /db_xref="taxon:37206"
BASE COUNT  501 a      607 c      854 g      516 t
ORIGIN
Query Match      1.3%; Score 71.2; DB 6; Length 2478;
Best Local Similarity 78.7%; Pred. No. 0.0028;
Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy      5205 TAAAGACTCTGGTAGTCCTCAAGCTTACCTGTTGAGCTGCTCGGTTGCATTCATTC 5264
        |||||
Db      108 TAAAGACTCTGGTAGTCCTCGCTGTACACGACGGGTCTCCGCGTTGCATTCATTC 49
        |||||
Qy      5265 CCAAGCAGCAAGAGGTGCGCAACTAGTAGCGCGCCCTCGGATACCA 5312
        |||||
Db      48 CCAAGCGGCAAGAGCAGTAGTACTGCTGCTCGCTCGCGGATACCA 1
        |||||

RESULT 15
HVScoat
LOCUS   HVScoat      2478 bp ss-RNA      linear      VRL 11-MAY-1995
DEFINITION Helicoverpa armigera Stunt Virus RNA2 coat protein (p71) and p17
            gene, complete cds.
ACCESSION U37299
VERSION   U37299.1 GI:609503
KEYWORDS coat protein; transfer RNA-like structure.
SOURCE   Helicoverpa armigera stunt virus
ORGANISM Helicoverpa armigera stunt virus
          Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
          unclassified Tetraviridae.
          1 (bases 1 to 2478)
REFERENCE Hanzlik,T.N., Dorian,S.J., Johnson,K.N., Brooks,E.M. and
AUTHORS Gordon,K.H.
TITLE     Sequence of RNA2 of the Helicoverpa armigera stunt virus
          (Tetraviridae) and bacterial expression of its genes
JOURNAL   U. Gen. Virol. 76 (Pt 4), 799-811 (1995)
MEDLINE   9049325
PUBMED    9049325
COMMENT   Original source text: Helicoverpa armigera stunt virus
          (individual isolate black mountain) cDNA to genomic RNA.
FEATURES
         source
         1..2478
            /organism="Helicoverpa armigera stunt virus"
            /mol_type="genomic RNA"
            /isolate="black mountain"
            /db_xref="taxon:37206"
            /clone="hasvna2"
          1..282
          1..50

5'UTR
stem_loop
1..50

```

```

/notes="stem loop which has hexamer sequence, gsuana, in
name place on a stem loop structure on RNA1; putative"
/function="unknown"
/gene="p756"
283..756
/gene="p17"
283..756
/gene="p17"
/function="unknown"
/notes="has PEST characteristics, makes tube-like
structures when expressed in bacteria, unstable when
expressed in baculovirus, start codon in poor context so
is poorly expressed if at all; putative"
/codon_start=1
/protein_id="AAC37884.1"
/db_xref="GI:609504"
/tranlation="MSEHTIASHITLPPGYTLALIPPEPAGMELEMRHSDLTVAE
PVTGSGAPFSPSPSEETNGVGRGKFLPTIISPLHKTRKALTPPSLPLTSLAC
PNSGIGPSPRSTPTIPSAOTSSTLQRYVLSIRASASTRSLTSSSSPSTOR"
366..2309
/gene="p71"
366..2309
/gene="p71"
/standard_name="HasV coat protein"
/notes="coat protein for HasV, forms T=4 capsids with 240
copies of protein per capsid, upon assembly into particle
is cleaved at N/P at residue positions 575-76"
/codon_start=1
/evidence=experimental
/product="coat protein"
/protein_id="AAC37885.1"
/db_xref="GI:609505"
/tranlation="MGDAGVASQRPNNRGTNRVRSANTVTYNGRRNRRTGRQVS
PPNFTFAAQDLAQSLDANTVTPPANISSMPEPRNAKGIIDISIGVFKLIDPA
GATESARAAVESEYKIPDLGVKFSVDALREIYNEBCVVDVDFLGRQMSLSIF
PMERTYVAAYANBEMSLDVVNDLIEMLNADMRVYVDSQMLFTDTTYVRI
RVLRPYVDVDPTEGIVRTVSDYRLTYKATLCEANMPTLVDOGFMGQVALPTSLP
QYVSEAYALHTLTPARPSAALATVYMAIIPGQTRPATTPAMEQSSGGYLTWRN
GTFPPAGSVSYVLPSPGALERVDYDNGSWTDFASAGTVFRQVAVDEVVANNPAG
GSAPFTFVRVPPSNAYTNVFRNTLLBTRPSRRLLELPMPPADFGQTVANNPKLBSL
LKETLCGYLVHSHKRNPFQLTLPASSFGVSPNNPGRYERDPLDYGIDSFPQMS
TAAVHPSLSHSCSIVTKYQGWEGVTNVTPPGQFHAAGLKNREBITLCLADLATL
TCYVPRATDNRAAASVAPAAAMLSVLKSRATSSIIISVGTAYGAAGSLAKPLGLM
SVPGKIAARVARRARRARRAN"
1497..1547
/gene="p71"
/notes="this region is found deleted in cDNA product after
reverse transcription by MuMolv RT or its derivative,
SuperScript I."
/evidence=experimental
2310..2478
/notes="putative"
2326..2478
/standard_name="viral tRNA-like structure"
/notes="tRNA-like structure without pseudo knot can be made
from this region, has valine anticodon, with 30 bases in
variable loop region; putative"
/function="unknown"

BASE COUNT  516 a      854 c      607 g      501 t
ORIGIN
Query Match      1.3%; Score 71.2; DB 14; Length 2478;
Best Local Similarity 78.7%; Pred. No. 0.0028;
Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy      5205 TAAAGACTCTGGTAGTCCTCAAGCTTACCTGTTGAGCTGCTCGGTTGCATTCATTC 5264
        |||||
Db      2371 TAAAGACTCTGGTAGTCCTCGCTGTACACGACGGGTCTCCGCGTTGCATTCATTC 2430
        |||||
Qy      5265 CCAAGCAGCAAGAGGTGCGCAACTAGTAGCGCGCCCTCGGATACCA 5312
        |||||
Db      2431 CCAAGCGGCAAGAGCAGTAGTACTGCTGCGCTCGCGGATACCA 2478
        |||||

```


Search completed: January 15, 2004, 11:22:55
Job time : 12472 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 14, 2004, 21:37:52 ; Search time 866 Seconds
(without alignments)
16558.216 Million cell updates/sec

Title: US-09-991-262-39
Perfect score: 5312
Sequence: 1 GTTTCGCTCCCTCCCGACGCG.....CGAGCGCTCCCTGGATACCA 5312

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 segs, 1349719017 residues
Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_19Jun03:*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5283.2	99.5	5312	15	AA058522
2	150	2.8	164	15	AA058524
3	76.4	1.4	6536	19	AAV04471
4	76.4	1.4	6536	19	AAV04471
5	71.2	1.3	172	15	AA058525
6	71.2	1.3	2478	15	AA058523
7	71.2	1.3	2478	19	AAV04471
8	50.8	1.0	1008	24	AA027039

9	50.2	0.9	1467	21	AAA99468
10	49.2	0.9	1556	24	AB090195
11	47.2	0.9	47981	22	AAE30757
12	47	0.9	27541	22	AAV17185
13	47	0.9	125401	22	AAV17185
14	46.8	0.9	4403765	22	AAI99683
15	45.6	0.9	5059	20	AAH84332
16	45.4	0.9	5871	24	AB878666
17	44.8	0.8	438	24	ABG91569
18	44.8	0.8	441529	22	AAI99682
19	44.6	0.8	4371	21	AAA99469
20	44.6	0.8	5241	21	AAA99467
21	44.4	0.8	44377	18	AAV78508
22	44.4	0.8	44377	18	AAV78508
23	44	0.8	77336	21	AAV14651
24	43.4	0.8	1044	15	AAQ66088
25	43.4	0.8	1891	24	ABV94243
26	43.4	0.8	1891	24	ABK84580
27	43.4	0.8	1891	24	ABV69280
28	43.4	0.8	1935	22	AAI58357
29	43.2	0.8	4257	19	AAV68520
30	43.2	0.8	4257	19	AAV10362
31	43.2	0.8	5692	22	AAV32248
32	42.8	0.8	747	22	AAH67991
33	42.8	0.8	1092	22	AAH44047
34	42.8	0.8	1092	24	ABV41287
35	42.8	0.8	1104	22	AAH78257
36	42.8	0.8	1104	22	ABV41287
37	42.8	0.8	2253	22	AAH67990
38	42.8	0.8	2466	25	ACA00124
39	42.8	0.8	6798	22	AAH44043
40	42.8	0.8	6798	22	AAH78258
41	42.8	0.8	6798	24	AAV31022
42	42.8	0.8	8077	24	AAV31028
43	42.8	0.8	349980	22	AAH68533
44	42.6	0.8	870	25	ACC00263
45	42.6	0.8	1468	25	ACC00262

ALIGNMENTS

AA058522	standard; DNA; 5312 BP.
AA058522;	
AC	AA058522;
XX	
DT	25-MAR-2003 (updated)
DT	12-SEP-1994 (first entry)
XX	
DE	Sequence of Heliothis armigera RNA 1 which encodes replicase, p1a,
DE	p1b and p14.
XX	
KM	HBSV, RNA 1; small RNA virus; p1a; p1b; p14; replicase; ss.
OS	
XX	Heliothis armigera stunt virus.
XX	
PH	Key
FT	37..5151
FT	Location/Qualifiers
FT	/tag= a
FT	/product= RBP1CAGE
FT	4218..4517
FT	/tag= b
FT	/product= p1a
FT	4518..4940
FT	/tag= c
FT	/product= p1b
FT	4944..5165
FT	/tag= d
FT	/product= p14
XX	
PN	MO9404660-A1.

Sakuranetin synthase
M. capulatus gene
Microspora meg
Streptomyces nours
Streptomyces nours
Mycobacterium tube
Stealth virus nucl
S. ghanensis DNA
M. capulatus gene
Mycobacterium tube
Sakuranetin synthase
Sakuranetin synthase
Platanolide synthase
Platanolide synthase
Nucleotide sequenc
HMDJUNX, a sample
Breast carcinoma r
Human cDNA differe
Prostate cancer re
Human polynucleot
The nucleotide seq
Infected cell prot
Streptomyces sp. C
C glutamicum codin
Streptomyces sp. C
Streptomyces IIP e
Nucleotide sequenc
Streptomyces IIP r
C glutamicum codin
C glutamicum codin
Streptomyces sp. C
Nucleotide sequenc
Streptomyces sp. C
Operon F DNA encod
C glutamicum codin
Antisense Maize CO
Maize COMT methyl

Db 1501 CGATCCACTCGCTCGAGAGACATATCTGCGACGTCGGCTGTCCTCCCTGAACAGAGTCGGC 1560
 Qy 1561 TTCTCTGCGCTCCCGCGTACCACTGCGCGGCTCTTTTCAAGACAGAGAAAGACTTGAAGTC 1620
 Db 1561 TTCTCTGCGCTCCCGCGTACCACTGCGCGGCTCTTTTCAAGACAGAGAAAGACTTGAAGTC 1620
 Qy 1621 CTTGCGAAGGTGAGCTGCTAACAAGAAAGTCCGGTACCTTCACTCTCTGTCGAGAG 1680
 Db 1621 CTTGCGAAGGTGAGCTGCTAACAAGAAAGTCCGGTACCTTCACTCTCTGTCGAGAG 1680
 Qy 1681 CCCCAGAGTTTGAAGCCGCACTTGTGAGCAGCAGCCGGGCTCACTCCCGAGTACCGC 1740
 Db 1681 CCCCAGAGTTTGAAGCCGCACTTGTGAGCAGCAGCCGGGCTCACTCCCGAGTACCGC 1740
 Qy 1741 GCGACCTTGGCAGGAGGTCTCAACACGAGCTCAAGACAGCTCAAGATACCTTGAAGAC 1800
 Db 1741 GCGACCTTGGCAGGAGGTCTCAACACGAGCTCAAGACAGCTCAAGATACCTTGAAGAC 1800
 Qy 1801 GCGCTCAAGACCATCGAGCGGCTCACTCTCCCACTGAGAGGCTCGAGATGTAAGAG 1860
 Db 1801 GCGCTCAAGACCATCGAGCGGCTCACTCTCCCACTGAGAGGCTCGAGATGTAAGAG 1860
 Qy 1861 GCGCCGCGCAGCAGCGGAGAGAGCGGCACTCTCACTGCGCGCTTGAAGCGCGGCGGT 1920
 Db 1861 GCGCCGCGCAGCAGCGGAGAGAGCGGCACTCTCACTGCGCGCTTGAAGCGCGGCGGT 1920
 Qy 1921 AAAGCATTTTACGTGGACCCACCAAGAGACTGAGAGAGGCTTATGAGACCGCGGATCAA 1980
 Db 1921 AAAGCATTTTACGTGGACCCACCAAGAGACTGAGAGAGGCTTATGAGACCGCGGATCAA 1980
 Qy 1981 CCGCGCTCGGCTGCGGCTAGAGCAATGTCGCGCTTGAAGTTTCCGTCGTCGACCGC 2040
 Db 1981 CCGCGCTCGGCTGCGGCTAGAGCAATGTCGCGCTTGAAGTTTCCGTCGTCGACCGC 2040
 Qy 2041 GAGGCGCGCTCTTCTGCTACCGTGGTTATCGACAGAGTCTTCACTGTCCTCGCTGTAC 2100
 Db 2041 GAGGCGCGCTCTTCTGCTACCGTGGTTATCGACAGAGTCTTCACTGTCCTCGCTGTAC 2100
 Qy 2101 GTGCGATCTGTCAGCGCTTGTCCCGAGCTCAAGATAGTCTTTGAGGAGCGTCCAC 2160
 Db 2101 GTGCGATCTGTCAGCGCTTGTCCCGAGCTCAAGATAGTCTTTGAGGAGCGTCCAC 2160
 Qy 2161 CAATTCGGGTTTATAGACTTCCAAAGGCAAGCGGAAATGCGCTGTGAGGAGCGTCCAC 2220
 Db 2161 CAATTCGGGTTTATAGACTTCCAAAGGCAAGCGGAAATGCGCTGTGAGGAGCGTCCAC 2220
 Qy 2221 GTTAAGCAGTCCGTCGCGGCACTTTCACCAAAACAAGCGCTGTCCGCGCAGCGTCTT 2280
 Db 2221 GTTAAGCAGTCCGTCGCGGCACTTTCACCAAAACAAGCGCTGTCCGCGCAGCGTCTT 2280
 Qy 2281 GCGACACAGTTTTCAGAGCTTGTACCCCGGAGTCACAACTCAAGGAGCGTCCGA 2340
 Db 2281 GCGACACAGTTTTCAGAGCTTGTACCCCGGAGTCACAACTCAAGGAGCGTCCGA 2340
 Qy 2341 TCCATTCAGCCAGCTGCGCCCAAGACTACCGCAAGCGGCGGCAAAAGCTCTGCTTACG 2400
 Db 2341 TCCATTCAGCCAGCTGCGCCCAAGACTACCGCAAGCGGCGGCAAAAGCTCTGCTTACG 2400
 Qy 2401 CAGAGAGAAAGTCCGCGCAAGCGGCTGAGGAGCGGATGACTGTGTCAGAGAGCGAGGA 2460
 Db 2401 CAGAGAGAAAGTCCGCGCAAGCGGCTGAGGAGCGGATGACTGTGTCAGAGAGCGAGGA 2460
 Qy 2461 CGCACTTTTGGTGTCTGATCTGATTAACAAGCGCTCCACAGCAGACAGAAAGCTCTCTC 2520
 Db 2461 CGCACTTTTGGTGTCTGATCTGATTAACAAGCGCTCCACAGCAGACAGAAAGCTCTCTC 2520
 Qy 2521 GCTGAGAGTGGCACTTCTGATCTGATTAACAAGCGCTCCACAGCAGACAGAAAGCTCTCTC 2580
 Db 2521 GCTGAGAGTGGCACTTCTGATCTGATTAACAAGCGCTCCACAGCAGACAGAAAGCTCTCTC 2580
 Qy 2581 GACCCGACAGGTGACATTTGAGAGCAACTCAACCATACCGGAAAGCGGAGTGTTCACA 2640
 Db 2581 GACCCGACAGGTGACATTTGAGAGCAACTCAACCATACCGGAAAGCGGAGTGTTCACA 2640

Qy 2641 GACATCCCTGCAACCCCTGAGATCAGACTGTCAACCGAGTGAAGGTTGACAGCGCAAC 2700
 Db 2641 GACATCCCTGCAACCCCTGAGATCAGACTGTCAACCGAGTGAAGGTTGACAGCGCAAC 2700
 Qy 2701 GAATGATGGAACAGATACCCCGCAGAGTGCACAGCGGCAAGAGCAATCTATCTGTC 2760
 Db 2701 GAATGATGGAACAGATACCCCGCAGAGTGCACAGCGGCAAGAGCAATCTATCTGTC 2760
 Qy 2761 CGCAAGAACTTCCGGGAGCAACCCGACTGTGCTGTGTCCGTTTGGGAAAGACCGGCTAC 2820
 Db 2761 CGCAAGAACTTCCGGGAGCAACCCGACTGTGCTGTGTCCGTTTGGGAAAGACCGGCTAC 2820
 Qy 2821 GAGGTGTTTGGCGGCTGTCGCAAAATGAAGTGAAGCTTGGCCGACCCGACGCAACCCG 2880
 Db 2821 GAGGTGTTTGGCGGCTGTCGCAAAATGAAGTGAAGCTTGGCCGACCCGACGCAACCCG 2880
 Qy 2881 AAGCGCATAGGAGGCTTCCAGAGAGGGGTACAGTGGTCAAGTCAACCAACGCGTCTAAC 2940
 Db 2881 AAGCGCATAGGAGGCTTCCAGAGAGGGGTACAGTGGTCAAGTCAACCAACGCGTCTAAC 2940
 Qy 2941 AAACACAGGCGCTCCAGAGGCTGTTCGCGTACACCAAGCGAAGCGCTGACCTGCGC 3000
 Db 2941 AAACACAGGCGCTCCAGAGGCTGTTCGCGTACACCAAGCGAAGCGCTGACCTGCGC 3000
 Qy 3001 CTACAGAGAGTAAAGAGAGAGCTCAACAGATGCTAACTCGCTTGAACCGCATTTGGGAC 3060
 Db 3001 CTACAGAGAGTAAAGAGAGAGCTCAACAGATGCTAACTCGCTTGAACCGCATTTGGGAC 3060
 Qy 3061 TGACCTGTCACTGAAGAGCGCCGCTGACCGAGCTGTCTTCCAGACCCACTCAAGTTAC 3120
 Db 3061 TGACCTGTCACTGAAGAGCGCCGCTGACCGAGCTGTCTTCCAGACCCACTCAAGTTAC 3120
 Qy 3121 CAACGCGCGGCACTGTCGAAGAGCTGTGAGAGCCGACAGACCCCTACATCTGTGACATA 3180
 Db 3121 CAACGCGCGGCACTGTCGAAGAGCTGTGAGAGCCGACAGACCCCTACATCTGTGACATA 3180
 Qy 3181 GACTTCCTTATGAAGACTCAGACGAAAGTGTGCGCAAGCCGATCAATACGAGGAGGTC 3240
 Db 3181 GACTTCCTTATGAAGACTCAGACGAAAGTGTGCGCAAGCCGATCAATACGAGGAGGTC 3240
 Qy 3241 GGGCAGGGAGTCCGCGCTCACTCAAGTCTCTCAACTTCTGCTGCGGCTTGAATACG 3300
 Db 3241 GGGCAGGGAGTCCGCGCTCACTCAAGTCTCTCAACTTCTGCTGCGGCTTGAATACG 3300
 Qy 3301 ATACTGAGAGATATCTCCGTACCGGAGCCGACAGTCCGGTACAGCAAGCTCTCC 3360
 Db 3301 ATACTGAGAGATATCTCCGTACCGGAGCCGACAGTCCGGTACAGCAAGCTCTCC 3360
 Qy 3361 GACGAAAGAGGCGCATGCTGCTGAAAGCGAAATCAATCAAGTCCCAAGCGGACGTTTC 3420
 Db 3361 GACGAAAGAGGCGCATGCTGCTGAAAGCGAAATCAATCAAGTCCCAAGCGGACGTTTC 3420
 Qy 3421 GTCTCGGCGAGCTGAGCCGAGTGTGACACGCGCCCAATTAACAGAGTGAAGTCTTTC 3480
 Db 3421 GTCTCGGCGAGCTGAGCCGAGTGTGACACGCGCCCAATTAACAGAGTGAAGTCTTTC 3480
 Qy 3481 GCGCGCTTTTGAAGCGCATGAGCAGCTGACGCTGCACTGATTAATTAATTAAGAGAGCG 3540
 Db 3481 GCGCGCTTTTGAAGCGCATGAGCAGCTGACGCTGCACTGATTAATTAATTAAGAGAGCG 3540
 Qy 3541 TGTGGAAGAGCGCACTTGGAGAGAGGTTAAGGTTAAGGTTAAGTGAAGTGAAGTCTGTC 3600
 Db 3541 TGTGGAAGAGCGCACTTGGAGAGAGGTTAAGGTTAAGGTTAAGTGAAGTGAAGTCTGTC 3600
 Qy 3601 GACTCCGGCGAGCTGAGAGCTTGGACGCTTGGCGCAACCAATCTTCTGCGCGGCTGATGTC 3660
 Db 3601 GACTCCGGCGAGCTGAGAGCTTGGACGCTTGGCGCAACCAATCTTCTGCGCGGCTGATGTC 3660
 Qy 3661 ACGCTCTTCCGCGGCGTCAAGTTCGAGAGCTTCAAAAGCGACGACTGCTCTGTGTGT 3720
 Db 3661 ACGCTCTTCCGCGGCGTCAAGTTCGAGAGCTTCAAAAGCGACGACTGCTCTGTGTGT 3720

PA	(CSTR -) COMMONWEALTH SCI & IND RES ORG.
PA	(PACI -) PACIFIC SEEDS PTY LTD.
PI	Christian PD, Gordon KHJ, Hanzlik TN;
DR	WPI, 1994-083180/10.
XX	
PT	Small RNA virus capable of infecting insect species, e.g.
PT	Heliothis - and transgenic plants contg. viral nucleic acid, for
PT	protection against insect pests
XX	
PS	Example; Figure 6; 183pp; English.
XX	
CC	H. armigera larvae were raised and viral RNA was extracted. The virus
CC	RNA were reverse transcribed into cDNA. Clone E3 represents 99.7%
CC	or RNA 1. hr236 contains about 88% or RNA 2. The genome of HasV
CC	has major differences that make it distinct from those of the
CC	nodaviruses, although it shares the characteristic of a bipartite
CC	genome with the nodaviridae. One difference is that the nodaviruses
CC	have a 3' blockage, whereas the HasV RNAs terminate in a
CC	distinctive secondary structure resembling a tRNA. AA058524 and
CC	AA058525 are RNA sequences which corresp. to the 3' UTRs of Figure
CC	1 (RNA 1) and Figure 2 (RNA 2) respectively. The FTRs indicate the
CC	tRNA-like structures.
CC	(Updated on 25-MAR-2003 to correct PN field.)
SO	
XX	Sequence 172 BP, 31 A, 48 C, 42 G, 37 U, 14 other;
XX	
Query Match	1.3%; Score 71.2; DB 15; Length 172;
Best Local Similarity	63.0%; Pred. No. 1,4e-07;
Matches	68; Conservative 17; Mismatches 23; Indels 0; Gaps 0;
OY	
5205	TAAAGACTCTGTGAGTCTCAACGTTACTCTGTGAGTCTGTGCGATTTCATTTC 5264
DB	65 UAAAGACUCUGUGAGUCCGCCGUGUACAAGACGAGUCUGCCGCGUUCGUAUCCAUUC 124
OY	
5265	CCAAAGCAGCAAAAGGTGGCGCACTGTACGCGCGCCCTGGGATTCGA 5312
DB	125 CCAAGCGGCAAGAAAGACGUAUGUACUCUGCUCGCCGGAUACCA 172
RESULT 6	
AA058523	
ID	AA058523 standard; DNA; 2478 BP.
XX	
AC	AA058523;
XX	
DT	25-MAR-2003 (updated)
DT	12-SEP-1994 (first entry)
XX	
DE	Sequence of Heliothis armigera RNA 2 which encodes P17 and the
DE	capsid proteins precursor P71.
XX	
KM	HasV; RNA 1; small RNA virus; P17; P71; ss.
XX	
OS	Heliothis armigera stunt virus.
XX	
FH	Key
FT	location/Qualifiers
FT	CDS 283..753
FT	/*tag= a
FT	/product= P17
FT	366..2309
FT	/*tag= b
FT	/product= P71
XX	
XX	WO9404660-A1.
XX	
PD	03-MAR-1994.
XX	
PF	13-AUG-1993; 93WO-AU00411.
XX	
RR	14-AUG-1992; 92AU-0004081.
RR	08-JUL-1993; 93US-0089372.

XX	(CSTR) COMMONWEALTH SCI & IND RES ORG.
PA	(PACT-) PACIFIC SEEDS PTY LTD.
XX	
PI	Christian PD, Gordon KHJ, Hanzlik TN;
XX	
DR	WPI, 1994-083180/10
XX	P-PSDB; AAR49661, AAR49662.
PT	Small RNA virus capable of infecting insect species, e.g.
PT	Heliothis - and transgenic plants contg. viral nucleic acid, for
PT	protection against insect pests
XX	
PS	Disclosure; Figure 2; 183pp; English.
XX	
CC	The inventors claim a virus comprising a genome hybridisable with
CC	the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are
CC	those given in Figs 1 and 2 of the specification. As isolated
CC	protein or polypeptide prepn. of the proteins or polypeptides
CC	derivable from the virus are also claimed.
CC	H. armigera larvae were raised and viral RNA was extracted. The virus
CC	RNA were reverse transcribed into cDNA. Clone hr236 contains about
CC	88% or RNA 2. RNA 2 encodes a protein of mol.wt.71,000(P71) which
CC	contains the peptide sequences corresp. to those determined from
CC	the two virus capsid proteins. This protein is therefore the
CC	precursor of these capsid proteins. In addition, another major
CC	translation product of apparent mol. wt. 24,000 is obtd. This
CC	protein is derived from a mol. wt. 17,000 reading frame overlapping
CC	the slab of the capsid protein gene. The Mr 24,000 protein (referred
CC	to as P17) may have a function in modifying or manipulating the growth
CC	characteristics or cell cycle of host-insected cells.
CC	(Updated on 25-MAR-2003 to correct PN field.)
CC	
XX	
Seq	Sequence 2478 BP; 516 A; 853 C; 608 G; 501 T; 0 other;
Query Match	1.3%; Score 71.2; DB 15; Length 2478;
Best Local Similarity	78.7%; Pred. No. 3.9e-07;
Matches	85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY	5205 TAAAGACTCTGTAGTCTCAAGCTTACTGTAGCTGTGCGGTTGGATTCCATTTC 5264
DB	2371 TAAAGACTCTGTAGTCTCCCGTGTTCACGACGCGGTCTGCCCGGTTCGATTCCATTTC 2430
QY	5265 CCAAGCAGCAAGAGGTGGCAACTAGTACGCGCCCTCGGATACCA 5312
DB	2431 CCAAGCGCAGGAAGAGAGTATTAGTCTGCGTCCCTCGGATACCA 2478
RESULT 7	
ID	AAT99117 standard; cDNA; 2478 BP.
AC	AAT99117;
XX	
DT	08-JUN-1998 (first entry)
XX	
DE	Helicoverpa armigera RNA2 encoding coat proteins p71 and p17.
XX	
KM	Vaccine; coat protein; p71; p17; insecticide; Ig-like domain; ds.
XX	
OS	Helicoverpa armigera.
XX	
Key	Location/Qualifiers
PH	283..756
FT	/*tag= a
FT	/product= p17 coat protein
FT	366..2309
FT	/*tag= b
FT	/product= p71 coat protein
XX	
PM	W09746666-A1.
XX	
PD	11-DEC-1997.

XX 02-JUN-1997; 97WO-AU00349.
 PF 31-MAY-1996; 96AU-0000234.
 XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
 XX Gordon KH, Hanzlik TN;
 PI WPI: 1998-042175/04.
 XX P-PSDB; AAM34534, AAM34535.
 DR Modified small RNA viruses and virus-like particles - have altered
 PT or substituted Ig-like domains to modify host cell tropism, useful
 PT as insecticides and in medicinal applications
 XX
 PS Disclosure: Figure 1; 41pp; English.
 CC The sequence is that of a cDNA encoding RNA2 of the viral genome which
 CC putatively codes for coat proteins p71 and p17. The p71 coat protein
 CC includes an Ig-like domain which can be used in the production of
 CC virus-like particles (VLP). The VLPS can be used in vaccines where the
 CC Ig-like domain has been altered so that the VLP presents a surface
 CC located antigen which is used to elicit an immune response in a host
 CC organism. They can also be used controlling the proliferation of a pest
 CC insect and potentially as medicinal delivery agents for cancer treatment
 CC and gene therapy.
 XX
 SQ Sequence 2478 BP; 516 A; 854 C; 607 G; 501 T; 0 other;

Query Match 1.3%; Score 71.2; DB 19; Length 2478;
 Best Local Similarity 78.7%; Pred. No. 3.9e-07;
 Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 5205 TAAAGACTGTGTGAGTCTCTCAACGTTACTGTGTGCGGTTGCATTCCATTC 5264
 DB 2371 TAAAGACTGTGTGAGTCTCTCTGTTACACGCGGTTGCGCGGTTGCATTCCATTC 2430
 OY 5265 CCAACGAGAAAGGTCGCACTAGAGGGCCCCCGGATGCA 5312
 DB 2431 CCAACGCGAGAGAGCGTAGTACTGTGCTGCTGCGGATGCA 2478

RESULT 8
 AAD27039
 ID AAD27039 standard; cDNA; 1008 BP.

AC AAD27039;
 XX
 DT 09-APR-2002 (first entry)
 XX
 DE Rice caffeic acid 3-O-methyltransferase (COMT) cDNA #1.
 XX
 KW Rice; caffeic acid 3-O-methyltransferase; COMT; caffeic acid;
 KW 5-hydroxyferulic acid O-methyl transferase; phenylpropanoid;
 KW transgenic plant; 5-hydroxyferulic acid; injury-repair mechanism;
 KW papermaking; host defect repair mechanism; lignin biosynthesis; ss.
 XX
 OS Oryza sativa.

XX Location/Qualifiers
 FH 67..983
 FT /tag= a
 FT /product= "Rice caffeic acid 3-O-methyltransferase"
 FT /note= "CDS does not include stop codon"
 FT /partial
 FT /transl_except= (pos:799..802, aa:Pro)
 FT /note= "This codon has an apparent one nucleotide
 FT insertion which alters the reading frame"
 FT /transl_except= (pos:854..857, aa:Gly)
 FT /note= "This codon has an apparent one nucleotide
 FT insertion which alters the reading frame"
 FT /transl_except= (pos:894..904, aa:Leu-Xaa-Pro-Asn)

FT /note= "Xaa corresponds to an unknown amino acid"
 FT /transl_except= (pos:935..942, aa:Tyr-Asp-Ala)
 FT /note= "Insertion of one base alters the reading frame"

XX US6329204-B1.
 XX 11-DEC-2001.
 XX 09-FEB-2000; 2000US-0500569.
 XX 10-FEB-1999; 99US-119587P.
 XX (DUPO) DU PONT DE NEMOURS & CO E I.
 XX Cahoon RE, Rafalekt JA, Shen JB;
 XX WPI: 2002-121138/16.
 XX P-PSDB; AAB16509.

DR New polynucleotide encoding caffeic acid 3-O-methyltransferase from
 PT rice, useful for studying lignin biosynthesis and in positive selection
 PT systems -
 XX
 PS Example 3; Page 23-24; 34pp; English.
 CC The invention relates to an isolated nucleic acid fragment encoding
 CC caffeic acid 3-O-methyltransferase (COMT) (also known as caffeic
 CC acid/5-hydroxyferulic acid O-methyl transferase) involved in
 CC phenylpropanoid metabolism. COMT is involved in the biosynthesis
 CC of the mono- or dimethoxylated lignin precursors. COMT DNA is used
 CC to produce transgenic plants or microbes that have altered levels
 CC of methylation for both caffeic and 5-hydroxyferulic acids, so can
 CC provide a method for positive selection. Since caffeic acid
 CC 3-O-methyltransferase is involved in lignin biosynthesis, the
 CC transgenic plants may have altered contents, or types, of lignin,
 CC so may be better suited for papermaking, and COMT DNA can also be
 CC used to study synthesis of lignin in plant cells, including its role
 CC in host defect- and injury-repair mechanisms. The present sequence
 CC is rice COMT cDNA.

SQ Sequence 1008 BP; 179 A; 343 C; 333 G; 152 T; 1 other;

Query Match 1.0%; Score 50.8; DB 24; Length 1008;
 Best Local Similarity 43.5%; Pred. No. 0.029;
 Matches 229; Conservative 0; Mismatches 298; Indels 0; Gaps 0;

OY 430 GTGCCACAGCGCGGCCGAGTTCAAGGCCGAGCTCACTGCTGCCAAGCGATTGCC 489
 DB 262 GCCGACAAGCTGCGTCCAAAGCGAACCAGCGGCGCGCATGTGACCGCATGCTC 321
 OY 490 TCCCGCACTTGTGCTGAGCGAAGTCCGCTTTGCGCGTTCAAAATCCGCGTTGAATT 549
 DB 322 CGCTGCTGCTCTTCAAACTGTCAAGTCCGAGTGAAGAGAGAGCGCCGACGCAAG 381
 OY 550 GCCAATCACTCCCTTATGACGTGACCTTAGAGAGCTGAGCAATGCTTTGAGAACCAC 609
 DB 382 CTCCTCCCGCGTGAAGCGCGCGCGCGGTGTGCAATGTGACACGCCCAAGAGACGCG 441
 OY 610 GGACTTCACATGATGTCGCGCTTCATGACATGACCAAGAGCTGCTTACATGACAAAC 669
 DB 442 GTCTCCATGAGCGCGCTCGCTCATGAACGAGCAAGAGTCTCATGAGAGCTGTAC 501
 OY 670 GTGTTAATGCGGACTGCGCTACCGCTTCCAGCTTATGAAGAGCTTATGCTGTGAAG 729
 DB 502 TACCTTAAGAGAGAGCTGAGNAAACGCGGCATCCCGTTCAACAAAGCGTACGAGTAGAG 561
 OY 730 GACTGCGATTCAGAGGAGGAGACCTCGCTTCCACTTCCGATGTGACTTATCAAC 789
 DB 562 GCGTTGATACACAGCAGCAGACCGCGCTTCAACCGGCTTCAACAGAGGCGATGAAG 621
 OY 790 GAGACCAAGAGCGGCGCATGAGAGGCTGAGCGCGCGCGGCTTCTTACTCCAGACGCGC 849
 DB 622 AACCACTCGGTATCATACCAAGAGGCTGAGCTGTACCTTACACCGGCTTTCAGCGCGC 681

Oy 850 GTGATTTTCTCCGCGACGACGATGGGGTGTGATCGTACTTACACGACTTCCACATGG 909
 Db 682 TCACACGCTGCTCAACGTCGCGCGGCGGCTGGGCGCCACTGTGCGCGCTGCTCCCGC 741
 Oy 910 CTGCGCTACTGCTGTGAGAACTACCCGACTCCGTTGGTTTCTC 956
 Db 742 CACCCGACATCCGGGGATCACTACGACTCCCGCAGTCATCTC 788

RESULT 9

AAA99468
ID AAA99468 standard; DNA; 1467 BP.

XX AAA99468;

XX 26-JAN-2001 (first entry)

XX Sakuranetin synthase related DNA sequence.

XX Sakuranetin synthetase; naringenin 7-O-methyltransferase; NMOT;

XX antibacterial; transgenic plant; ds.

XX Oryza sativa.

XX WO200052174-A1.

XX 08-SEP-2000.

XX 03-MAR-2000; 2000MO-JP01306.

XX 04-MAR-1999; 99JP-0057748.

XX (NISC-) JAPAN SCI & TECHNOLOGY CORP.

XX Kodama O;

XX WPI; 2000-572186/53.

PT Sakuranetin synthetase gene encoding protein with naringenin
 7-O-methyltransferase activity in rice cells and promoter activity,
 used to derive sakuranetin from naringenin easily in plants to impart
 antibacterial characteristics -

XX Claim 5; Page 53-54; 57pp; Japanese.

CC This invention relates to a sakuranetin synthetase gene encoding a
 CC protein with naringenin 7-O-methyltransferase (NMOT) activity and
 CC promoter activity. The invention includes a recombinant vector including
 CC the sakuranetin synthetase DNA and host cells transformed with the
 CC vector. The sakuranetin synthetase protein has antibacterial activity and
 CC is used to produce transgenic plants with antibacterial characteristics.
 CC The present sequence represents DNA related to the sakuranetin synthetase
 CC gene of the invention.

XX Sequence 1467 BP; 279 A; 458 C; 452 G; 278 T; 0 other;

XX Query Match 0.9%; Score 50.2; DB 21; Length 1467;

XX Best Local Similarity 43.5%; Pred. No. 0.048;

XX Matches 229; Conservative 0; Mismatches 298; Indels 0; Gaps 0;

XX Oy 430 GTGGCCACAGGCGCGCGGATTCAAGCCGACCTCTACTGCTCCGCAACGGCATTC 489

XX Db 278 GCCGACAGCTGCGCTCAAGCGAACCCGCGCGCGACGACATGATGACCGCATCTC 337

XX Oy 490 TCCGCGACCTTCTGCGTGAAGAGTGGCTCTTGCCGCTTCAATGCGCGTTGAATT 549

XX Db 338 CGCTGCTCGCTCTCAACGCTGTCAAGTGCAGATGAGAGGCGCGCGACGCGAG 397

XX Oy 550 GCCAATCACTCTCTATGAGAGTACCCCTAGAGAGAGCTGGCCAAATGCTTTAGAACAC 609

XX Db 398 CTCTCCGCGCTACGCGCGCGCGGATGCAAGTGGCTGACGCCCAAGAGACGCGC 457

Oy 610 GGAATTACATGCTCCGCGGCTTCATGACATGTCAGAAAGCTGCTCAATGACAAAC 669
 Db 458 GTCTCATAGCCGCCCTTCGCCCTCATGAAACAGAGAAAGTCTCATGAGAGCTGTAC 517
 Oy 670 GTGGTTAATGCGGAGCTCGGCTACCGCTTCCAGTTATGAAAGCCTATAGCTGTGAAG 729
 Db 518 TACCTTAAGGACGACATCTCTGAGACGGCGCATCCGTTCAACAAAGCGTACGGGATGACG 577
 Oy 730 GACTGGCAATTCAGAGGGGGGAGCTCCGTTCTCACTTCCCTGAGTTGACATTCAAC 789
 Db 578 GCGTTGAGTACACGACGACGACGCGCGCTTCAACGCGCTTCAACGAGGCGCATGAAAG 637
 Oy 790 GAGAGCCAAGAGGGGCGCATCGAGAGCTGAGCGCGCGCTCTTCACTCCAGACGCGCC 849
 Db 638 AACCACTCCGCTATCATACCAAGAAAGCTGCTGACCTTCAACCGCTTCAACGCGCGCC 697
 Oy 850 GTGATTTTCTCCGCGACGACGACGCTGGGGTGTATGCTTACACGACTTCCACATGG 909
 Db 698 TCACGCTGCTGACGCTCGGCGGCGGCGGTGGGCGCATGTGGCGCGCGCTCTCCCGC 757
 Oy 910 CTGCGCTACTGCTGTGAGAACTACCCGACTCCGTTGGTTTCTC 956
 Db 758 CACCCGACATCCGGGGATCACTACGACTCCCGCAGCTCATCTC 804

RESULT 10

ABQ90195
ID ABQ90195 standard; DNA; 1656 BP.

XX ABQ90195;

XX 01-OCT-2002 (first entry)

XX M. capsulatus gene #180 for DNA array.

XX Micro array; gene; ds; differential expression; gene expression.

XX Methyllococcus capsulatus.

XX WO200255655-A2.

XX 18-JUL-2002.

XX 14-JAN-2002; 2002MO-NO000019.

XX 12-JAN-2001; 2001NO-0000235.

XX 12-JAN-2001; 2001NO-0000239.

XX (UNIF-) UNIFOB STIFTELSEN UNIV BERGEN.

XX (TIGR-) TIGR.

XX Birkeland NK, Bidhammer I, Jonassen I, Jensen HB, Ilen T;

XX Lillhaug JR, Lossius I, Risen JA, Frazer CM, Durkin AS;

XX Salzberg SL;

XX WPI; 2002-557818/59.

XX Novel DNA array useful for determining differential expression of

XX Methyllococcus capsulatus genes, comprises polynucleotides or

XX oligonucleotides representative for a selective number of Methyllococcus

XX capsulatus genes -

XX Claim 19; Page 139; 678pp; English.

CC The invention relates to a novel DNA array giving a representation of a
 CC number of Methyllococcus capsulatus genes. The method of the invention is
 CC useful for determination of the differential expression of the genes of
 CC M. capsulatus, and for studying gene expression on a genomic scale and in
 CC gene expression assays of M. capsulatus genes. The sequences shown in
 CC ABQ9016-ABQ9185 represent M. capsulatus genes for use in arrays of the
 CC invention.

XX Sequence 1656 BP; 328 A; 572 C; 494 G; 262 T; 0 other;


```

FT      synthase 2"
FT      /note= "polyketide synthase, encodes AAB82213"
FT      22957..24237
FT      /*tag= x
FT      /gene= "megA11"
FT      /function= "KS3"
FT      24544..25581
FT      /*tag= y
FT      /gene= "megA11"
FT      /function= "AT3"
FT      26230..26733
FT      /*tag= z
FT      /gene= "megA11"
FT      /function= "KR3 (inactive)"
FT      26998..27258
FT      /*tag= aa
FT      /gene= "megA11"
FT      /function= "ACP3"
FT      27393..28590
FT      /*tag= ab
FT      /gene= "megA11"
FT      /function= "KS4"
FT      28897..29931
FT      /*tag= ac
FT      /gene= "megA11"
FT      /function= "AT4"
FT      29953..30477
FT      /*tag= ad
FT      /gene= "megA11"
FT      /function= "DH4"
FT      31396..32244
FT      /*tag= ae
FT      /gene= "megA11"
FT      /function= "ER4"
FT      32257..32799
FT      /*tag= af
FT      /gene= "megA11"
FT      /function= "KR4"
FT      33052..33312
FT      /*tag= ag
FT      /gene= "megA11"
FT      /function= "ACP4"
FT      33666..43271
FT      /*tag= ah
FT      /gene= "megA11"
FT      /product= "megalomicin 6-deoxyerythronolide B
FT      synthase 3"
FT      /note= "polyketide synthase; encodes AAB82214"
FT      22957..24237
FT      /*tag= ai
FT      /gene= "megA11"
FT      /function= "KS5"
FT      24544..25581
FT      /*tag= aj
FT      /gene= "megA11"
FT      /function= "AT5"
FT      26230..26733
FT      /*tag= ak
FT      /gene= "megA11"
FT      /function= "KR5"
FT      26998..27258
FT      /*tag= al
FT      /gene= "megA11"
FT      /function= "ACP5"
FT      27393..28590
FT      /*tag= am
FT      /gene= "megA11"
FT      /function= "KS6"
FT      28897..29931
FT      /*tag= an
FT      /gene= "megA11"
FT      /function= "AT6"
FT      29953..30477
FT      misc_feature

```

```

FT      /*tag= ao
FT      /gene= "megA11"
FT      /function= "KR6"
FT      31396..32244
FT      /*tag= ap
FT      misc_feature

Query Match      0.9%; Score 47.2; DB 22; Length 47981;
Best Local Similarity 46.2%; Pred. No. 1;
Matches 157; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

QY      1651  CCGGTACCTTCACCTCCTCTGTGAGAGAGCCCAAGTTTCAGCCCGGACTTGCGCAC 1710
DB      30391  CCGGAGCCACCGCCCTGACGTTGACCGACCACTACGACACCTCCGCGAAGTGGGTAC 30450

QY      1711  GCGACCGCGGCTCACTCCCGAGTACCGGCGCACTTGAGGAGAGTTCACACCGAC 1770
DB      30451  GAGTACGGGCGCGGCTTCAGAGGCGTGGCGCGCGTGGAGACGCGGAGTGTCTAC 30510

QY      1771  GTCAGAGCTCAAGATCAACCTTCAGAAAGCCCTCAAGACATGACGGGCTCACCCTC 1830
DB      30511  GCGAGAGTCTCTCAACGCGCTGAGAGAGGTACGCTTCAGCCCGGTGCTGAC 30570

QY      1831  TCCCAATCAGAGCGCTCGAGATGTACGAGGCGCCGCAAGCGGCAAGAGCGGACCC 1890
DB      30571  GCGGTGCGCCAGACCTTCGCGCTGACGACGCGCGGCGCACTGCGGTACGCGGTGGCGACCCCGCC 30630

QY      1891  CTCATGCGCGCCCTTGAAGCGCGCGCGGTAAAGCACTTACGTGACCCACGAGAA 1950
DB      30631  CGGCGGCTCACCTTCGACGCGCACCGGCGCGCACTGCGGTACGCGGTGGTGGACCCCGCC 30690

QY      1951  CTGAGAGGCGTATGAGCGGCGGATCAACGCGCGTCCG 1990
DB      30691  GAGCGGACCGGTTGGCGCTTGGCGGTACCGACCGACCG 30730

RESULT 12
AAD17185
ID      AAD17185 standard; DNA; 27541 BP.
XX      AC      AAD17185;
XX      DT      29-NOV-2001 (first entry)
XX      DE      Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX      KW      Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;
XX      KW      antifungal; antibiotic; nys2; ds.
XX      OS      Streptomyces noursei.
XX      FH      Key
XX      FT      CDS
XX      Location/Qualifiers
XX      /*tag= a
XX      /product= "NysF protein"
XX      /note= "CDS does not include start codon"
XX      complement (1275..3092)
XX      /*tag= b
XX      /product= "NysG protein"
XX      complement (3070..4824)
XX      /*tag= c
XX      /note= "CDS does not include start codon"
XX      /product= "NysH protein"
XX      5122..6156
XX      /*tag= d
XX      /product= "NysD3 protein"
XX      6338..27541
XX      /*tag= e
XX      /product= "NysI partial protein"
XX      /note= "CDS does not include stop codon"
XX      WO200159126-A2.
XX      16-AUG-2001.

```

```

XX 08-FEB-2001; 2001WO-GB00509.
PE
XX 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UNO-) UNIV NORGES TEKNIISK NATURVITENSKAPLIGE.
PA (SINTF) SINTER STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEMSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVRIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zorchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR P-PSDB; AAE10138, AAE10139, AAE10140, AAE10141, AAE10142.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 2; Page 151-166; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nys2 DNA of nystatin PKS gene cluster.
XX
SQ Sequence 27541 BP; 3517 A; 10766 C; 9529 G; 3729 T; 0 other;

```

```

Query Match 0.9%; Score 47; DB 22; Length 27541;
Best Local Similarity 47.5%; Pred. No. 0.92;
Matches 140; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

```

```

OY 1633 GAGCTGCTCAACGAAAGTCCGGTACCTTCCTCCGCGGAGAGAGCCCAAGTTTC 1692
DB 9330 GACCGCGTCGAGAACTCACTCGCGGACCTCTCTGCTGCCGGAACAGGGCGCGTC 9379
OY 1693 GACCGCGCTTGCGCAAGCGGACCGGCGCTCACTCCCGAGTACCGGCGCACTTGAG 1752
DB 9380 CAGGTCAGTTGTGATCGGCAACCGCGAGTGTCCGGTCGCCGACCGTCAAGCTCCAC 9439
OY 1753 GCAAGTCTCAACACCGAGCTCAAGAGCTCAAGATCACTCGAGAGAGCGCTCAAGACC 1812
DB 9440 GCCCGCGCCGACACCGGAGCAACCCCTGGAACCGGCCACGCCGCGCTCTCTCACCC 9499
OY 1813 ATCGAGGGGCTCACCTCTCCCACTCAAGAGCGCTCAAGATGATAGAGGCGCGCGCAGGC 1872
DB 9500 ACCGCGGAGCGCTCCCGCAGTCTCCCGGCTTGTCTCGAGCAAGGGGCGACCCCTCTGCG 9559
OY 1873 AGCGGCAAGAGCGGAGCGCTCACTCCCGCGCTTGAGGCGCGCGGCGTAAAGCAC 1927
DB 9560 GCGGACCCCGCAACCGCGCGCTGAGCGGCGGCGCAGTGGCCCGCGCGGCGCGCAAC 9614

```

```

RESULT 13
AADI1786
ID AADI1786 standard; DNA; 125401 BP.
XX
AC AADI17186;
XX
DT 29-NOV-2001 (first entry)
XX
DE Streptomyces noursei nystatin PKS gene cluster DNA.
XX
KW Polyketide synthase; PKS; macrolide; nystatin; PKS gene cluster;

```

```

KW antifungal; antibiotic; de.
XX
OS Streptomyces noursei.
XX
FH Key Location/Qualifiers
FT CDS 6337..34771
FT /tag= a
FT /product= "NysI complete protein"
FT 34792..51099
FT /tag= b
FT /product= "NysJ protein"
FT 51155..57355
FT /tag= c
FT /product= "NysK protein"
FT 57503..58687
FT /tag= d
FT /product= "NysL protein"
FT complement (58786..58980)
FT CDS
FT /tag= e
FT /product= "NysM protein"
FT /note= "CDS does not include start codon"
FT complement (59045..60241)
FT /tag= f
FT /product= "NysN protein"
FT /note= "CDS does not include start codon"
FT complement (60238..61296)
FT /tag= g
FT /product= "NysD2 complete protein"
FT 120628..121308
FT /tag= h
FT /product= "NysR4 (long) protein"
XX
EN W0200159126-A2.
XX
PD 16-AUG-2001.
XX
PE 08-FEB-2001; 2001WO-GB00509.
XX
XX 08-FEB-2000; 2000GB-0002840.
PR 10-APR-2000; 2000GB-0008786.
PR 14-APR-2000; 2000GB-0009387.
XX
PA (UNO-) UNIV NORGES TEKNIISK NATURVITENSKAPLIGE.
PA (SINTF) SINTER STIFTELSEN IND TEK FORSK.
PA (ALPH-) ALPHARMA AS.
PA (SINV-) SINVENT AS.
PA (DZIE/) DZIEGLEMSKA H.
PA (ZOTC/) ZOTCHEV S B.
PA (SEKU/) SEKUROVA O N.
PA (FJAE/) FJAEVRIK E.
PA (BRAU/) BRAUTASET T.
PA (STRO/) STROM A R.
XX
PI Zorchev SB, Sekurova ON, Fjaervik E, Brautaset T, Strom AR;
PI Valla S, Ellingsen TE, Sletta H, Gulliksen O;
XX
DR WPI; 2001-557614/62.
DR P-PSDB; AAE10143, AAE10144, AAE10145, AAE10146, AAE10147, AAE10148,
DR AAE10149, AAE10150.
XX
PT New nystatin polyketide synthase polynucleotides and polypeptides,
PT useful as antibiotics and antifungals -
XX
PS Claim 1; Page 188-254; 266pp; English.
XX
CC The present invention relates to the cloning and sequencing of the gene
CC cluster encoding a modular type I polyketide synthase (PKS) enzyme
CC involved in the biosynthesis of the macrocyclic antibiotic nystatin.
CC The nystatin PKS is useful as antifungal antibiotics. The present
CC sequence is a Streptomyces noursei nystatin PKS gene cluster DNA.
XX
SQ Sequence 125401 BP; 15664 A; 49692 C; 42871 G; 17174 T; 0 other;

```

Query Match 0.9%; Score 47; DB 22; Length 125401;
 Best Local Similarity 47.5%; Pred. No. 1.7; Indels 0; Gaps 0;
 Matches 140; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 1633 GGCTGTACAAAGCAAGTCCGATACCTTCACATCTCTGTGAGAGAGCCCAAGTTTC 1692
 Db 9320 GACCGGTGAGAGAACTCAACCTTCGCGCACCCTCTGTGTCGCCAAGAGGCGGCTC 9379
 QY 1693 GACGCGACTTGTGCAAGCGACCGGAGCTCACTCCCGAGTACCGGCGACCTTGAC 1752
 Db 9380 CAGGTTCAGTTGTGATCGGCAACCCCGAGTGTCCGATGCGCGCACCCGCAAGTCCAC 9439
 QY 1753 GCAGGTCTAAGACCGACGTCAGCAAGTCAATACCTTCGAGAACCGGCTCAAGACC 1812
 Db 9440 GCCCGCCGACACCGGCGAGACACCCCTGAGACGCGCCACCGGCGTCTCAC 9499
 QY 1813 ATGACGAGGCTCAACCTTCCTCCCAAGTACAGGAGCTGATACGAGGCGCGGAGG 1872
 Db 9500 ACCGCGGAGCGCTCCCGGCAAGTCCGAGTCCGAGAGAGGCGGCAACCCCTCGCC 9559
 QY 1873 AGCGGCAAGACGGGCAACCTTCATGCGCGGCTTGAGGCGCGGCGGTAAAGCAC 1927
 Db 9560 GCGGACCCCGACCCCGCTCGACGCGGCGCAGTGGCCCGGCGGCGCGAAC 9614

RESULT 14
 AA19683/c
 ID AA19683 standard; DNA; 4403765 BP.
 AC AA19683;
 DT 15-JUN-2002 (first entry)
 XX
 XX Mycobacterium tuberculosis strain H37Rv genome SEQ ID NO 2.
 DE Mycobacterium tuberculosis; strain H37Rv; strain CDC 1551; genome;
 KM variation; epidemiology; patient treatment; epidemic monitoring; ds.
 XX
 OS Mycobacterium tuberculosis.
 XX
 PN US6294328-B1.
 XX
 PD 25-SEP-2001.
 XX
 PF 24-JUN-1998; 98US-0103840.
 XX
 PR 24-JUN-1998; 98US-0103840.
 XX
 PA (GENO-) INST GENOMIC RES.
 XX
 XX Fleischmann RD, White OR, Fraser CM, Venter JC;
 PI Fleischmann RD, White OR, Fraser CM, Venter JC;
 XX
 DR WPI; 2001-647261/74.
 XX
 PT Evaluating strain variation of Mycobacterium tuberculosis, comprises
 PT determining the nucleotide sequence of the strain at positions in the
 PT genome corresponding to positions where M. tuberculosis strains CDC
 PT 1551 and H37Rv differ -
 XX
 PS Claim 4; SEQ ID NO 2; 3pp + Sequence Listing; English.
 CC
 CC The invention relates to evaluating strain variation within and between
 CC different populations of the tuberculosis bacterial pathogen,
 CC Mycobacterium tuberculosis or related Mycobacterium by determining the
 CC nucleotide sequence of the first strain at positions in the complete
 CC sequence of the genome that correspond to positions that differ in the
 CC nucleotide sequences of M. tuberculosis strains CDC 1551 (AA19683) and
 CC H37Rv (AA19682). The method is useful for evaluating strain variation of
 CC M. tuberculosis and has valuable application in the fields of
 CC tuberculosis genetics, epidemiology, patient treatment and epidemic
 CC monitoring.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from USPTO

CC at seqdata.uspto.gov/sequence.html?docId=6294328B1.
 XX
 XX Sequence 4403765 BP; 757105 A; 1447799 C; 1441301 G; 757371 T; 189 other;
 SQ

Query Match 0.9%; Score 46.8; DB 22; Length 4403765;
 Best Local Similarity 50.4%; Pred. No. 7.4; Indels 0; Gaps 0;
 Matches 114; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 345 CGCCCCGTGCGAGATATACGAGGTGACCAAGTACGAGCGGCGAGCGGCTCGGAC 404
 Db 3922657 CGACGCGGTGCGCGGGAATACGCGCTTGCCTCCGCGCACCGCGGCGCACCTTGCCGG 3922598
 QY 405 CATTAAGGCTTAAGATCTAATCCCTGCGCACAGGCGGCGCGGCTTCAAGCCGACGC 464
 Db 3922597 CATTCGCGCTTGGCGCGGCGGCGGCGGCTTACCGGCGGCGGCAATTCGCGCTG 3922538
 QY 465 CTCACGTCTGCGCAAGGCAATTCCTCCGCGACCTTCTGAGTGAAGAGTGGGCTTTG 524
 Db 3922537 CGCGCGGCGCGCACCGTTCGCGCGCTCTACCTTTGGCGCGGCGGCGGCGGCGGCT 3922478
 QY 525 CGGCTCAATGCGCGGCTTGAATTCGCAATCACTTCCTATGAC 570
 Db 3922477 CTGTGCAAGTGCCTCGGTATGCGGCTGCGCGGCGGCGCAATGCC 3922432

RESULT 15
 AAX84332/c
 ID AAX84332 standard; DNA; 5059 BP.
 XX
 XX AAX84332;
 AC
 DT 08-SEP-1999 (first entry)
 XX
 XX Stealth virus nucleic acid clone, SEQ ID NO: 24.
 DE
 XX
 KM Stealth virus; detection; diagnosis; infection; ss.
 XX
 OS Stealth virus.
 XX
 XX
 FH Key Location/Qualifiers
 FT misc_difference 3605
 FT /tag= a
 FT /note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 FT
 FT misc_difference 3610
 FT /tag= b
 FT /note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 FT
 FT misc_difference 3615
 FT /tag= c
 FT /note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 FT
 FT misc_difference 3630
 FT /tag= d
 FT /note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 FT
 FT misc_difference 3631
 FT /tag= e
 FT /note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 FT
 FT misc_difference 3638
 FT /tag= f
 FT /note= "this nucleotide is represented as a * in the
 FT specification, and is included to maintain the
 FT base numbering given in the specification"
 FT
 FT misc_difference 3641
 FT /tag= g
 FT /note= "this nucleotide is represented as a * in the

[illegible]

```
FT      misc_difference 3860      base numbering given in the specification"
FT      /*tag= ak
FT      /note= "this nucleotide is represented as a * in the
FT      specification, and is included to maintain the
FT      base numbering given in the specification"
FT      misc_difference 3863
FT      /*tag= a1

Query Match      0.9%; Score 45.6; DB 20; Length 5059;
Best Local Similarity 15.1%; Pred. No. 1.1;
Matches 70; Conservative 177; Mismatches 209; Indels 9; Gaps 1;

QY      3763 GAACGTTACAGACCAAAATTTGAAGCTGAGGTGCAAAATCTGCCGTACATCGGA 3822
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      482 RAVATKASDNSSYDANATCYNMHSSTKSDGSDCYDASSNHTGCMKKDAMYANGS 423

QY      3823 CTCCTGCTCCCGCTGAGCAGGTGCTCTGACCCCTGTCAAGAGCGCTCAAGATATT 3882
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      422 TRTKSYBNTAYGGNAKKNKSAGHGANCSTYBDSYBRNGHKKNTKDDAMNAGNAN 363

QY      3883 GGGCGCTGTACACAGCGAATCTCTTACTCCAAAGTACGTGAGAGCTGTGAGACATC 3942
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      362 HNGRITDADKGRAYATKASYSTKTYHTKSKASRNNKDAYGNTGSNAGNYBSRTS 303

QY      3943 ACCAAGGCTGAGTGAGCGCCGCTACCAAGCCTCTGTGCCACATGTCAGCATGCTAC 4002
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      302 NSGHAGTVENHAGNCTKTHNKKYTGAGHYRVTVTNGHTSNYSNNKYTGRTKYTAT 243

QY      4003 TACAATTAGCGCGCGGA-----CTCGCGCGGTACATCATCGACGCTGTTGTCGC 4053
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      242 YNCAHYHYHADACNHYGRTKNAGNSYBAGTNTSKAKRSSSYBRNCYSTYDAAYYSGH 183

QY      4054 TTGGGCGCGGCACTTCCCGTTTGAACAACCTGCGCGTGTGCGGCCATGTGCAGGCA 4113
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      182 AGNCONRYASHRVSYBNCVATGHTYNTGNSYNKSSGGSSTSNAGNSYBAHNSYHNTNCG 123

QY      4114 CCGAGCGCTTACAGACAGCAGTATCCGGCTAACGTGCGGCATGTCCTTGACACGTC 4173
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      122 YASKASRSNNKYTGDNAGNVTKTSYBTCKARGNTKSRMNNKNCNTGTNTNATATHTDA 63

QY      4174 TTGAGCGCGCGCGCGCGCGCGCGCGCGAGGTTTCGTGCGACA 4218
      | : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      62 TASVVVNNYBNNNNVVBNBNNBBBNNBBBNNBBBNNVV 18
```

Search completed: January 15, 2004, 07:54:53
Job time : 881 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 15, 2004, 05:48:31 ; Search time 6688 Seconds
(without alignment(s))
19304.034 Million cell updates/sec

Title: US-09-991-262-39

Perfect score: 5312

Sequence: 1 GTTCGCTCCCGGACG.....CGCGCCCGCTGGATACCA 5312

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estrov:*
6: em_estrpl:*
7: em_estro:*
8: em_estro:*
9: gb_est1:*
10: gb_est2:*
11: gb_est3:*
12: gb_est4:*
13: gb_est5:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gsa_hum:*
18: em_gsa_inv:*
19: em_gsa_pin:*
20: em_gsa_vrt:*
21: em_gsa_fun:*
22: em_gsa_mam:*
23: em_gsa_mus:*
24: em_gsa_pro:*
25: em_gsa_rtd:*
26: em_gsa_vhl:*
27: em_gsa_vhl:*
28: gb_gsa1:*
29: gb_gsa2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	55.2	1.0	1201	13	BX381961 BX381961
2	53.6	1.0	417	124	BM442184 EBAN01.SO
3	53.2	1.0	629	14	CA596350 wpa1c.DK0
4	53	1.0	524	12	B0246178 B0246178

5	52.6	1.0	565	13	B0753689	B0753689 EBAN01.SO
6	52.2	1.0	235	12	BM442316	BM442316 EBAN01.SO
7	52	1.0	608	14	BG605165	BG605165 WHE2328.C
8	51.6	1.0	627	10	CA741702	CA741702 w1a1c.DK0
9	51.4	1.0	659	12	B0247417	B0247417 B0247417
10	50.8	1.0	925	29	CNS0091P	AL053013 Drosophila
11	50.4	0.9	925	29	CNS0091P	AL053013 Drosophila
12	50.2	0.9	792	14	CB645456	CB645456 OSUTRE07C
13	49.6	0.9	885	13	BK425603	BK425603 BX425603
14	49.6	0.9	1101	29	CNS0175Y	AL108460 Drosophila
15	49.4	0.9	1011	23	BQ961334	BQ961334 AGENCOURT
16	49.2	0.9	816	10	BF263602	BF263602 HV.CBA000
17	49.2	0.9	920	12	BG837702	BG837702 zml0.D190
18	49	0.9	519	14	CA708438	CA708438 wdr2c.DK0
19	48.8	0.9	291	9	AV428623	AV428623 AV428623
20	48.8	0.9	412	13	BQ245036	BQ245036 TAB15032A
21	48.8	0.9	466	12	B0248671	B0248671 B0248671
22	48.8	0.9	554	6	AL813432	AL813432 Tricium
23	48.8	0.9	582	10	BG273981	BG273981 WHE2232.A
24	48.8	0.9	606	12	B0247799	B0247799 B0247799
25	48.8	0.9	648	12	B0257566	B0257566 B0257566
26	48.8	0.9	686	12	B0293750	B0293750 B0293750
27	48.8	0.9	713	13	BQ804210	BQ804210 WHE3551.G
28	48.4	0.9	410	14	CA594759	CA594759 wpa1c.DK0
29	47.6	0.9	424	10	BE606576	BE606576 WHE9001.C
30	47.6	0.9	778	14	CB652764	CB652764 OSUNEC03B
31	47.6	0.9	1170	11	AY108821	AY108821 Zea mays
32	47.4	0.9	588	14	CB873044	CB873044 HCL0101Y
33	47.2	0.9	462	13	B0983444	B0983444 HA29M03T
34	47.2	0.9	465	12	BM441172	BM441172 EBED02.SQ
35	47.2	0.9	554	12	B0478971	B0478971 B0478971
36	47.2	0.9	586	14	CA019759	CA019759 HM13A10T
37	47.2	0.9	587	13	B0939346	B0939346 HM05B10T
38	47.2	0.9	605	13	B0994738	B0994738 HM08A17T
39	47.2	0.9	622	14	CB882761	CB882761 HL02M11W
40	47.2	0.9	652	13	B0607000	B0607000 BRY.2879
41	47.2	0.9	655	13	B0465368	B0465368 HU03G11T
42	47.2	0.9	896	10	AM982441	AM982441 HVSMB000
43	47.2	0.9	899	29	BZ553726	BZ553726 pac81-60
44	47	0.9	425	10	BE590730	BE590730 WHE0857.G
45	47	0.9	1201	13	BX332497	BX332497 BX332497

ALIGNMENTS

RESULT 1
LOCUS BX381961/c 1201 bp mRNA linear EST 08-MAY-2003
DEFINITION BX381961 Homo sapiens PLACENTA COT 25-NORMALIZED Homo sapiens cDNA
clone CS0D1072YF05 3-PRIME, mRNA sequence.
ACCESSION BX381961
VERSION BX381961
KEYWORDS BX381961.1 GI:30453007
SOURCE EST.
ORGANISM Homo sapiens (human)

REFERENCE
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1201)

TITLE Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
JOURNAL Full-length cDNA libraries and normalization
COMMENT Unpublished
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr Web: www.genoscope.cns.fr

LIBRARY Library was constructed by Life Technologies, a division of
Invitrogen. Contact: Feng Liang Email: fliang@lifestech.com URL:
http://fulllength.invitrogen.com/InvitrogenCorporation 1600
Faraday Avenue Genoscope sequence ID: CS0D1072CC03NP1.
Location/Qualifiers
1..1201
/organism="Homo sapiens"

FEATURES
source

```

/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="CS0D1072YF05"
/issue_type="PLACENTA COT 25-NORMALIZED"
/issue_type="PLACENTA COT 25-NORMALIZED"
/clone_1ib="Homo sapiens PLACENTA COT 25-NotI-oligo (dT)
/notes="1st strand cDNA was primed with a NotI-oligo (dT)
primer. Five prime end enriched, double-strand cDNA was
digested with Not I and cloned into the Not I and EcoR V
sites of the pCMVSPORT 6 vector. Library was normalized."
BASE COUNT
ORIGIN
95 a      191 c      115 g      55 t      745 others

```

Query Match	1.0%: Score 55.2: DB 13; Length 1201;
Best Local Similarity	6.4%: Pred. No. 0.048;
Matches	52; Conservative 275; Mismatches 485; Indels 0; Gaps 0;
Qy	1986 GTCGCGCTCGGCTACGCAACATGTCGCGCTTTCGATTCTCGCTGTCGACCGCGAGG 2045
Dy	1186 GCGCCGCGCC 1127
Qy	2046 CGCCCGCTTTCGCTACCGCGCTTATCGAGAGTGCCTATGTTCCGCGCTGCGTACGTCG 2105
Dy	1126 GKKGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 1067
Qy	2106 GATCGTCGACGCGCTTGTCCCGAGCTCAAGAATGCTCTTGTAGGGAGCGTCCAAAT 2165
Dy	1066 KKK 1007
Qy	2166 CGGCTTTATACCTTCCAAAGCACAAGCGCAACATGCGCGCTCTTCCGCACTCGTTAA 2225
Dy	1006 MGGGMMKVVMMGGKGVKGVKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 947
Qy	2226 GACGTGCGCGTGGCGGCACTTTCACCAACCAAGCGCTGTCGCGCGGAGTGGTGGCAC 2285
Dy	946 KTTMMMBBKK 887
Qy	2286 CACGTTTTTCCAGAGCTTGTACCCCGGTGCACACACTCAGAGGTGCGTGCATCCAT 2345
Dy	886 KKGKKTGKK 827
Qy	2346 CAGCGACGCGCGCCAGACTACCGCAACGCGAGCGGCAACGCTCTGCTTCCACAGAA 2405
Dy	826 MNNKKKKKKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGGKGG 767
Qy	2406 GGAAGAGTCGCGGCCACGCGGCTGAGGGCGCGATGACTGTGACGAAGCGCAGACGCA 2465
Dy	766 MNNNNNNKKNNKKVAMNNNNNNNGKNNKKNNNNNNNNNNNNNNNNNNNNNNNNNN 707
Qy	2466 TTTTGGCTGTGTCATTTGCATTTACACGCGTCCACAGAGAGCAGAACTCTGCGCTGA 2525
Dy	706 NNN 647
Qy	2526 GAAGTGCGACCTTCTAAGTCGCGATCAGCGCGCACACACACTGTATCCTCGGACCC 2585
Dy	646 MNN 587
Qy	2586 GACAGGTGACATTGAGAGACAACTCAACATAGCGGAAAGCGGAGGTGTTTACAGACAT 2644
Dy	586 NNN 527
Qy	2646 CCCTGACCCCTCGAGATCAGACTGTCTAAACCGAGTGAAGAGTGCAGCGCAAGAAAT 2705
Dy	526 KNNKK 467
Qy	2706 GATGCAACAGATACCCCGCGAGATGCCACGCGCGCACGGAGCAATCATCTGCTCGCAA 2765
Dy	466 NNN 407
Qy	2766 GAACTTGGGAGCAACCGGACTGTGCTGTG 2797
Dy	406 GNNNNNNNTGTGKMMNNNNNTGTGNNCGATTGK 375

```

RESULT 2
BM442184
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES
SOURCE

BM442184          417 bp      mRNA      linear      EST 23-JUL-2002
Eban01_S0002_H11_R anther, yellow stage, no treatment, cv Optic,
Eban01_Hordeum vulgare subsp. vulgare cDNA clone Eban01_S0002_H11
5', mRNA sequence.
BM442184
BM442184.2  GI:21931275
EST.
Hordeum vulgare subsp. vulgare
Hordeum vulgare subsp. vulgare
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidaeae
; Triticeae; Hordeum.
1 (bases 1 to 417)
Hedley, P., Liu, H., Caldwell, D., McCallum, N., Mudie, S., Cardle, L.,
Ramsay, L., Machray, G., Marshall, D.F.M. and Maugh, R.
Development of Barley Transcription Resources
Unpublished
On Feb 1, 2002 this sequence version replaced gi:18472959.
Contact: Maugh R, Marshall DF
Genome Dynamics/Computational Biology
Scottish Crop Research Institute
Invergowrie, Dundee, DD2 5DA, Scotland, UK
Tel.: 00 44 1382 562731
Fax: 00 44 1382 562426
Email: est@scri.sari.ac.uk
All sequence has a Phred quality score of 20 or over
Seq primer: M13 reverse.
location/Qualifiers
1..417

```

BASE COUNT	85 a	148 c	115 g	69 t
ORIGIN				
Query Match		1.0%;	Score 53.6;	DB 12; Length 417;
Beet Local Similarity		57.9%;	Fred. No. 0.077;	
Matches 95; Conservative		0;	Mismatches 69; Indels	0; Gaps 0;
Qy	1733	AGTACCGGCGACCTTGCAGGCGAGGTCTCAACACGAGTCAAGAGCTCAAGATCACCC		1792
Db	253	ACTTCGACCCCATTCAGCTGGGGGAGATGAAAGACCAAGCTGCCAGCTCGAAGCATGC		312
Qy	1793	TCGAGAACGCGCCTCAAGACCATCGACGGGCTTCACCTTCTCCCGATCAGAGGCTCTGAGA		1855
Db	313	TCTCCAAACACGGACCAAGAACATGCGGCACCATCTGACTCTCCAAAGACCAAGACGGCCAGA		372
Qy	1853	TGTACGAGGCGGCCCGCCAGGACGGGCAAGATGAGGGGCAACCTTCATC		1896
Db	373	TGTACTTGCGCTTCGGCGGGCGCCGACAGTGTTCGCCAAGCTGTC		416
RESULT 3				
CA596350				
LOCUS	CA596350	629 bp	mRNA	linear
DEFINITION	wpa1c.pK012.k24 wpa1c Triticum aestivum cDNA clone wpa1c.pK012.k24			
	5' end; mRNA sequence.			
ACCESSION	CA596350			

VERSION	KEYWORDS	ORGANISM	REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT	FEATURES	source
CAS96350.1	GI:25147773	Triticum aestivum (bread wheat)	EST.						
		Triticum aestivum							
		Triticum aestivum							
		Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae							
		; Triticeae; Triticum.							
		1 (bases 1 to 629)							
		Tingey, S.V., Moore, G., Griffiths, S., Powell, W., Wolters, P., Dolan							
		, M., Hailey, C., Miao, G., Caraher, N., and Hanley, W.K.							
		Dupont Wheat cDNA sequence in collaboration with the John Innes							
		Center 1							
		Unpublished							
		Contact: Scott V. Tingey							
		Crop Genetics							
		E. I. Dupont de Nemours and Company							
		1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA							
		Tel: 302-631-2602							
		Fax: 302-631-2607							
		Email: Scott.V.Tingey@usa.dupont.com							
		Seq primer: M13.							
		Location/Qualifiers							
		1..629							
		/organism="Triticum aestivum"							
		/mol_type="mRNA"							
		/db_xref="taxon:4565"							
		/clone="wpaic.pK012.k24"							
		/issue_type="anchors"							
		/lab_host="DH10B"							
		/clone_lib="wpaic"							
		/note="Vector: pBluescript SK+, Site_1: EcoRI, Site_2:							
		XhoI; Wheat [Triticum aestivum] pre-meiotic anthers JIC"							
		140 a 212 c 186 g 86 t 5 others							
		ORIGIN							
		Query Match							
		Best Local Similarity 48.5%; Pred. No. 0.11; Length 629;							
		Matches 142; Conservative 0; Mismatches 151; Indels 0; Gaps 0;							
		Y 1758 TCTCAACACCGCAGCTCAAGCATCAACCTCGAGAACGCTCAAGACCATCGA 1817							
		Db 330 TCAACACCTTTCAACAAGAGCTCCGCGTGGCCCCGAGAGCATCCCGTCTGCTACCGA 389							
		Y 1818 CGGGCTACCTCTTCCCGAGTCAAGGCGCTCGAAGTGAAGAGGAGGAGGAGGAGGAGG 1877							
		Db 390 GGGCCCCCTCAACCCCAAGGCAACCGTGAAGATGATGATGATGATGATGATGATGATGAT 449							
		Y 1878 CAACACGGGACCGCTCATCGCGCGCTTGAAGCGCGCGGCGGTAAAGACATTACGTGGC 1937							
		Db 450 CAACACCGCGCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 509							
		Y 1938 ACCACACGAGAACTGAGAGAGGCTATGAGACCGGCGGATCAAAACCGCGCTCGCTCGGC 1997							
		Db 510 AACACCGGATATCGTCTCGACTCCGAGAGCGGTGTAGCCACACGAGGCGCATATTAAGN 569							
		Y 1998 TAGCAACATGTGCGCCCTTGGGATTTCTCGTGTGACACCGCGGAGGAGGCGCC 2050							
		Db 570 NMGTTAGCGCGTCCCGCCACCGCATCTCTCGTCTGACCTCCCGGACCGGAC 622							
		RESULT 4							
		BU246178							
		LOCUS							
		DEFINITION							
		ACCESSION							
		VERSION							
		KEYWORDS							
		SOURCE							
		ORGANISM							
		BU246178							
		524 bp							
		mRNA							
		linear							
		EST 05-APR-2002							
		BU246178 Y. Ogihara unpublished cDNA library, Wh_f Triticum							
		aestivum cDNA clone wh21n02 5', mRNA sequence.							
		BU246178							
		GI:20057927							
		EST.							
		Triticum aestivum (bread wheat)							

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1 (bases 1 to 524)	Ogihara, Y. and Mureai, K.	Expressed genes in Triticum aestivum	Unpublished	Contact: Tadao Shin-1 Center For Genetic Resource Information National Institute of Genetics 111 Yata, Mishima, Shizuoka 411-8540, Japan Tel: 81-559-81-6856 Fax: 81-559-81-6855 Email: tehin@genie.nig.ac.jp. location/Qualifiers
FEATURES				
SOURCE				
ORGANISM				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
RESULT 5				
BQ753689				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				
LOCUS				
DEFINITION				
Accession				
Version				
Keywords				


```

/lab host="E. coli SOLR"
/clone_lib="Wheat pre-anthesis spike cDNA library"
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;
Site_1: EcoRI; Site_2: XhoI; Plants were grown in the
greenhouse. Whole spike with awns trimmed, white, green
and yellow anther were collected and total RNA, and
poly(A) RNA were prepared, a cDNA library was made, and
the cDNA clones were in vivo excised to give phagescript
phagemids in the T3 clone lab (Choi, Close, Penson) at
the University of California, Riverside. Plasmid DNA
preparations and DNA sequencing were performed in the OD
Anderson lab (all other authors)."
BASE COUNT      116 a      221 c      179 g      91 t      1 others
ORIGIN
Query Match      1.0%; Score 52; DB 10; Length 608;
Best Local Similarity 48.5%; Pred. No. 0.21;
Matches 142; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

OY      1758 TCTCAACACCGACGTACAGCAGCTCAAGATCAACCTCGAGGAAGCCCTCAAGACCATCGA 1817
DB      264 TCACACCTTTCTACAAAGACCTCGCTGAGCCCCGAGAGACATCCGTCCTGCTACCGA 323
OY      1818 CGGGCTCACCTCTCTCCCACTGACAGGCTCGAGATGTACAGAGGCGCCGCAAGGACGGG 1877
DB      324 GCGGCCCCCTCAACCCCAAGGCCAAGCGTGAAGATGACCCAGATCATGTTGAGACCTT 383
OY      1878 CAAGACGGGACCCCTCATGTCGCGCCTTGAGCGCGGCGGTAAAGCACTTACGTGCG 1937
DB      384 CAACACCCCCCGCATGTACGTGCGCATCCAGGCGGTGCTCTCCCTACGCGCTCGGGGG 443
OY      1938 ACCCAACAGAACTAGAGAGGCTATGACCGCGCGATCAACCGCCGCTCGCCTCGGC 1997
DB      444 AACCCACGGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 503
OY      1998 TAGCGAACAATGTGCGCCTTGCGATTTCTCGGTGTCGACCGCGGAGGCGGCC 2050
DB      504 NCGGTACGGGCTGCGCCGACGCTCTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556

RESULT 8
LOCUS      CA741702      627 bp      mRNA      linear      EST 26-NOV-2002
DEFINITION      wialc.pk003.122 wialc Triticum aestivum cDNA clone wialc.pk003.122
ACCESSION      CA741702
VERSION      CA741702.1 GI:25557525
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
REFERENCE      Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
Tingley,S.V., Powell,M., Woltere,P., Dolan,M., Hainey,C., Yuan,Z.,
Miao,G., Caraher,N. and Hanafey,M.K.
TITLE      Dupont Wheat cDNA Sequence
JOURNAL      Unpublished
COMMENT      Contact: Scott V. Tingley
Crop Genetics
E.I. Dupont de Nemours and Company
1 Innovation Way, P.O. Box 6104, Newark, DE 19714-6104, USA
Tel: 302-631-2602
Fax: 302-631-2607
Email: Scott.V.Tingley@USA.dupont.com
Seg primer: M13.
FEATURES
Location/Qualifiers
1..627
/organism="Triticum aestivum"
/mol_type="mRNA"
/db_xref="taxon:4565"
/clone="wialc.pk003.122"
/tissue_type="anthers"

```

```

/lab host="DH10B"
/clone_lib="wialc"
/note="Vector: pBluescript SK+, Site_1: EcoRI; Site_2:
XhoI; Wheat (Triticum aestivum, Hi Line) Immature anthers"
BASE COUNT      129 a      209 c      184 g      89 t      16 others
ORIGIN
Query Match      1.0%; Score 51.6; DB 14; Length 627;
Best Local Similarity 47.6%; Pred. No. 0.27;
Matches 138; Conservative 0; Mismatches 152; Indels 0; Gaps 0;

OY      1758 TCTCAACACCGACGTACAGCAGCTCAAGATCAACCTCGAGGAAGCCCTCAAGACCATCGA 1817
DB      306 TCACACCTTTCTACAAAGACCTCGGTCGCCCCGANNAGATCCGCTCTCTCTACCGA 365
OY      1818 CGGGCTCACCTCTCTCCCACTGACAGGCTCGAGATGTACAGAGGCGCCGCAAGGACGGG 1877
DB      366 GCGGCCCCCTCAACCCCAAGGCCAAGCGTGAAGATGACCCAGATCATGTTGAGACCTT 425
OY      1878 CAAGACGGGACCCCTCATGTCGCGCCTTGAGAGCCGCGGCGGTAAAGCACTTACGTGCG 1937
DB      426 CAACACCCCCCGCATGTACGTGCGCATCCAGGCGGTGCTCTCTCTACGCGCTCGGGGG 485
OY      1938 ACCCAACAGAACTAGAGAGGCTATGACCGGCGGATCAACCGCGCTCGCCTCGGC 1997
DB      486 AACCAACGGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 545
OY      1998 TAGCGAACAATGTGCGCCTTGCGATTTCTCGGTGTCGACCGCGGAGGCGG 2047
DB      546 NNNGTACGGGCTGCGCCCAAGCGCATCTCGTCTCGANNCTGCGCGAGCGG 595

RESULT 9
LOCUS      BU247417      659 bp      mRNA      linear      EST 05-APR-2002
DEFINITION      BU247417 Y. Ogihara unpublished cDNA library, Wh_f Triticum
aestivum cDNA clone whf2c01 5', mRNA sequence.
ACCESSION      BU247417
VERSION      BU247417.1 GI:20058635
KEYWORDS      EST.
SOURCE      Triticum aestivum (bread wheat)
ORGANISM      Triticum aestivum
REFERENCE      Bukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
AUTHORS      Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae
Ogihara,Y. and Murai,K.
TITLE      Expressed genes in Triticum aestivum
JOURNAL      Unpublished
COMMENT      Contact: Tadasu Shin-I
Center For Genetic Resource Information
National Institute of Genetics
111 Yata, Mishima, Shizuoka 411-8540, Japan
Tel: 81-559-81-6856
Fax: 81-559-81-6855
Email: tshin@genes.nig.ac.jp.
FEATURES
Location/Qualifiers
1..659
/organism="Triticum aestivum"
/mol_type="mRNA"
/cultivar="Chinese Spring"
/db_xref="taxon:4565"
/clone="whf2c01"
/tissue_type="spike at flowering date"
/dev_stage="Feekes' scale 10.5.1"
/clone_lib="Y. Ogihara unpublished cDNA library, Wh_f"

```

```

BASE COUNT      135 a      248 c      172 g      104 t
ORIGIN
Query Match      1.0%; Score 51.4; DB 12; Length 659;
Best Local Similarity 48.5%; Pred. No. 0.3;
Matches 142; Conservative 0; Mismatches 151; Indels 0; Gaps 0;

```

QY 1758 TCTCAACCGAGCTCAAGAGCTCAAGATCACCCTCGAAGACGCTCAAGACATCGA 1817
 Db 103 TCACACCTTCTACAGAGCTCCGCGTGGCCCCGAGAGACATCCCTCTGTCTACCGA 162
 QY 1818 CGGAGCTACCTCTCCCACTCAAGAGCTCGAAGATGATGAGAGGCCCGGACGAGCGG 1877
 Db 163 GGGGCCCCCTAACCCCAAGGCCAAGCTGAGAGATGACCAATGATGTTCTGAGACTT 222
 QY 1878 CAAGACGGGACGCTCATCCGCGCTTGAAGCGCGCGGCGGTAAAGCATTTACGTGCG 1937
 Db 223 CAACACCCCGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 282
 QY 1938 ACCCACCAGAACTGAGAGAGCTATGACCGGCGGATCAACCGCGCTCGCTGCGC 1997
 Db 283 AACCAACCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 342
 QY 1998 TAGGCAACATGTGCGCTTGGGATTCCTGCGTGGCGACCGCGAGGCGCGCC 2050
 Db 343 GGGGATCGCGCTGCGCGCGCGCATCTCTGATGATGATGATGATGATGATGATGATG 395

RESULT 10
 CNS0091P/c 925 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T&T3 end of BAC #
 DEFINITION BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL053013.1 GI:4934461
 VERSION GSS.
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 925)

REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

COMMENT
 - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila>

FEATURES
 source
 1..925
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR19D16"
 /clone_1lb="RPCI-98"
 /note="end : T&T3"
 Location/Qualifiers

BASE COUNT 120 a 61 c 61 g 172 t 511 others

Query Match 1.0%; Score 50.8; DB 29; Length 925;
 Best Local Similarity 13.5%; Pred. No. 0.48;
 Matches 53; Conservative 173; Mismatches 166; Indels 0; Gaps 0;

QY 1583 CTGCCCCGCTTTACAGACGAGGAAGCTTGAGTCTTTCGCAAGCTGCTCTCA 1642

Db 925 CBBSCSCSCSBS 866
 QY 1643 ACGAACGTCGGTACCTTCACTCTCTGTGAGAGAGCCCAAGTTTCAAGCCGACT 1702
 Db 865 SSGTSSACVCKNAASBS 806
 QY 1703 TGTGACACCGACCGCGGCTCACTCCCGAATACCGGCGCACTTTCGAGGCGACTCA 1762
 Db 805 GGAASHSSSSACBS 746
 QY 1763 ACACGACGCTCAAGCTCAAGATCACTCGAGAACCCCTCAAGACATCAAGCGGC 1822
 Db 745 ASAGSVSSASBS 686
 QY 1823 TCACCTTCTCCAGTCAAGAGCTTCGAGATGATGACGAGCCCGCAGCAGCGCAAG 1882
 Db 685 TSMSCGCTSSASMSARSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBSBS 626
 QY 1883 CGGCGACCTCATGCGCGCGCTTGAAGCGCGCGGTAAAGCATTTAGTGCACCCA 1942
 Db 625 SMSGGGSGSVASBS 566
 QY 1943 CCAGAACTGAGAGAGCTATGACCGCGCG 1974
 Db 565 SGCACRCSGSSAABAAAASCVAAACGMMCGSK 534

RESULT 11
 CNS0091P 925 bp DNA linear GSS 03-JUN-1999
 LOCUS Drosophila melanogaster genome survey sequence T&T3 end of BAC #
 DEFINITION BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION AL053013.1 GI:4934461
 VERSION GSS.
 KEYWORDS Drosophila melanogaster (fruit fly)
 SOURCE Drosophila melanogaster
 ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 1 (bases 1 to 925)

REFERENCE
 AUTHORS Direct Submission
 TITLE Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
 JOURNAL BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr

COMMENT
 - Web : www.genoscope.cns.fr
 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP).
 The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org/TheBDGP/Drosophila>

FEATURES
 source
 1..925
 /organism="Drosophila melanogaster"
 /mol_type="genomic DNA"
 /db_xref="taxon:7227"
 /clone="BACR19D16"
 /clone_1lb="RPCI-98"
 /note="end : T&T3"
 Location/Qualifiers

BASE COUNT 120 a 61 c 61 g 172 t 511 others

Query Match 1.0%; Score 50.8; DB 29; Length 925;
 Best Local Similarity 13.5%; Pred. No. 0.48;
 Matches 53; Conservative 173; Mismatches 166; Indels 0; Gaps 0;

[illegible]

RESULT 12	LOCUS	DEFINITION
CB645456	CB645456	
	792 bp	linear mRNA EST 08-APR-2003
	OSNNE07C10.f	OSNNE07C10.f
	clone OSNNE07C10.5	clone OSNNE07C10.5
		CDNA

ACCESSION	CB645456
VERSION	CB645456.1
	GI:29640447

KEYWORDS	EST.
SOURCE	Oryza sativa (japonica cultivar-group)
ORGANISM	Oryza sativa (japonica cultivar-group)

REFERENCE
1 (bases 1 to 792)

AUTHORS Jantaauriyart,C., Lu,G., Gowda,M., Hatfield,J., Zhou,B., Mazur,E.,
Kudrna,D., Dean,R., Soderlund,C., Wing,R. and Wang,G.
TITLE Large-scale identification of ESTs involved in the interaction
between rice and Magnaporthe grisea
JOURNAL Unpublished
COMMENT Contact: Rod Wing

Arizona Genomics Institute
University of Arizona
Biological Sciences West, 448A, P.O. Box 210086, Tucson, AZ
85721-0086, USA
Tel: 520 626 3967
Fax: 520 621 9288
Email: <http://genome.arizona.edu>
PCR Primers
FORWARD: gta aaa cga cgg cca gtg
BACKWARD: gga aac agc tat gac cat g
plate: 07 row: C column: 10
Seq primer: gta aaa cga cgg cca gtg.

```

FEATURES
source
location/Qualifiers
1..792
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="mRNA"
/cultivar="Nipponbare"
/db_xref="taxon:39947"
/clone="OSJNB07C10"
/tissue_type="leaf"
/dev_stage="3 week"
/lab_host="DH10B"
/clone_id="OSJNB"
/note="Vector: pBluescript II KS +; Site_1: EcoRI; Site_2

```

	XhoI; 24 hrs after immunization with Rice Blast (Che 86061)			
)"			
BASE COUNT	142 a	270 c	241 g	138 t
ORIGIN	1 others			

Query Match	0.94	Score	50.2	DB	14	Length	792
Best Local Similarity	46.94	Pred.	No. 0	63			
Matches 157, Conservative	0	Mismatches	178	Indels	0	Gaps	0

622 GTCGCGGTTTCATGCACATGCAGAAAGAGCTCTTACATGGACAAGTGGTTAATGCC 681

DB 35 GGCCTCGCCCTCAGAACCAAGACAMGGTCTCATGAGAGAGCTGGTACTACCTTAAGGAC 94

Db 95 GCAGTCTCGGACGGCGGCATCCCGTTCAACAAGCGCTACGGGATGACGGCGCTTCGAGTAC 154

Oy 742 CAGGGGGGGGACCTCCGCTCCACTTCCCTGAGTGGCTTCACTCAACGAGGCCAAGAG 801
 Db 155 CACGGCACGGAGCCGCCGCTTCAACCGCGTCTTCAACGAGGGGCATGAAACCACTCCGTC 214

802 CGGCGCATCGAGGCGTGGCCGCGCGCGCTCTACTCCAGACGGCGCGTCAATTTCTCC 861

By	862	GGCGACGACGACTGGGGTGATGCGTACTTACACGACTTCCACACATGAGCTCGGCTACCTTA	921
Db	215	ATCATCACCAAGAGCTCGACCTCTACACCGGCTTCGAGCGCGGCTCCACCGTCGTC	274

Db 275 GACGTCCGCGCGCGCGCGCGCCACTGTGAGCCGCCCGTCGTCTCCCGCCACCCGCACTC 334

Qy	922	CTGGTGAAGAACTACCCCACTCCGTTTGGTTTCTC	956
Db	335	CGGGGANTCAACTACGACCTCTCCCACTGATCTTC	369

RESULT 13			
EX425603/c	885 bp	linear	EST 15-MAY-2003
LOCUS	EX425603 Homo sapiens	NEUROBLASTOMA	Homo sapiens cDNA clone
DEFINITION	C10BB0222A07 3-PRIME, mRNA sequence.		

ACCESSION	EX425603
VERSION	EX425603.1
KEYWORDS	GI:30770486
SOURCE	EST.
	<i>Homo sapiens</i> (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 885)
AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.
TITLE Full-length cDNA libraries and normalization

Unpublished
Journal
Comment
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: segret@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by Life Technologies, a division of
Invitrogen Contact : Peng Liang Email : fliang@lifetech.com URL :
<http://lifetech.invitrogen.com/> Invitrogen Corporation 1600
Paradise Avenue Genoscope sequence ID : CLO850222A07F1.

```

FEATURES
    source
        Location/Qualifiers
            1..885
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /clone="Cl08B0222A07"
                /cissue="NEUROBLASTOMA"
                /clone_lba="Homo sapiens NEUROBLASTOMA"
                /note="Vector: pCMVSPORT 6; 1st strand cDNA was primed
                    with a NotI-oligo(dT) primer. Five prime end enriched,
                    double-strand cDNA was digested with Not I and cloned into
                    the Not I and EcoRV sites of the pCMVSPORT 6 vector.
                    Library was not normalized."
BASE COUNT
    15 a      23 c      69 g      154 t      624 others
ORIGIN

```

[illegible]

```

FEATURES
    source
        pbeloBAC11.
            location/Qualifiers
                1..1101
                    /organism="Drosophila melanogaster"
                    /mol_type="genomic DNA"
                    /db_xref="taxon:7227"
                    /clone="BACN37L08"
                    /clone_1lb="DrosBAC"
                    /plasmid="pbeloBAC11"
                    /note="end : SP6"

BASE COUNT      254 a       176 c       160 g       152 t       359 others
ORIGIN
Query Match          0.9%; Score 49.6; DB 29; Length 1101;
Best Local Similarity 15.1%; Pred. No. 0.98;
Matches 54; Conservative 171; Mismatches 132; Indels 1; Gaps 1;

QY      1611 CAAAGAACATCGGTACTCTTCACCTCCTCTGAG AGAGGCCCAAGTTTGACCCG 1639
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      697 CACAACOCGCACATMACCCSSASBSASSSSSCASTSSARGMWSSCASGSAGSAGAC 756
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      1700 ACTTGTGGCACGCGACCGCGGCTCATCTCCCAGTAGTACCGCCACCTTCAGCAGATC 1759
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      757 ASAGAGSGSGRGSGGCCCLASGVCACACSSASCSASMSICVTSGSSCCASGSCCVSSSC 816
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      1760 TCAAACACCGACGTCAAGGAGCTCAAGTCAACCTTCGAAGAAGCCCTCAAGACCATCAGC 1819
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      817 AVASASVASVSVMKVASAVASCVAVASQMASGAVSSCRSVASVASAASVSBSVSSSS 876
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      1820 GGCTCACCCCTCCGCCAGTCAAGGAGCTTCGAGATGTACAGAGGCCCGCCAGGACGCGCA 1879
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      877 SSVVSAANAASSSSSASBASMAVAALAAAVASVSVAASVSSSSCSSSSASVASV 936
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      1880 AGACGAGCACCCCTCATCGCCGCCCTTGAAGCGCGGCGGTAAAGCATTACGTGAC 1939
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      937 AASASVSSSSSSSVSTSSASVSASVSASVASMSAVSSASSSSSSVSVVAVALAASAAA 996
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
QY      1940 CCACCAAGAACTAGAGAGGCTTAGACCGCGGCGATCAACCGCCCTCGCCTCGGC 1997
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||
DB      997 AAAAAAASSSSASAIVAVASSSSSSSASSSSSSSSVSSSSSVSSSCSVSVSSSS 1054
           |||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||::|||

RESULT 15
BO961334/c
LOCUS      BO961334               1011 bp      mRNA      EST 21-AUG-2002
DEFINITION AGENCOURT 8931666 NIH_MGC_40 Homo sapiens CDNA clone IMAGE:6484147
VERSION     5'', mRNA sequence.
ACCESSION   BO961334
VERSION     BO961334
KEYWORDS    BO961334.1 GI:22376812
SOURCE      EST.
ORGANISM    Homo sapiens (human)
EXTRACT     Homo sapiens
Eukaryotic Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1011)
NIH-MGC http://mgc.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished
Contact: Robert Straubeberg, Ph.D.
Email: csapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLML)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLML at:
http://lmlncl.lnl.gov
Plate: LILCM2669 row: n column: 20
High quality sequence stop: 269.
location/Qualifiers
    1..1011
        /organism="Homo sapiens"
        /mol_type="mRNA"
```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 15, 2004, 07:06:25 ; Search time 233 Seconds

(without alignments)
10062.778 Million cell updates/sec

Title: US-09-991-262-39

Perfect score: 5312

Sequence: 1 GTTCGCTCCCGCGACGG.....CGGCGCCCTGGATACCA 5312

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued Patents NA:*
1: /cgn2_6/pdata/2/ina/5B_COMB.seq:*
2: /cgn2_6/pdata/2/ina/5A_COMB.seq:*
3: /cgn2_6/pdata/2/ina/6A_COMB.seq:*
4: /cgn2_6/pdata/2/ina/6B_COMB.seq:*
5: /cgn2_6/pdata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/pdata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5312	100.0	5312	3	US-08-485-355B-39 Sequence 39, Appl
2	5312	100.0	5312	3	US-08-485-355B-41 Sequence 41, Appl
3	5312	100.0	5312	3	US-08-485-355B-43 Sequence 43, Appl
4	5206	98.0	5368	3	US-08-485-355B-45 Sequence 45, Appl
5	76.4	1.4	6534	3	US-09-194-613-4 Sequence 4, Appl
6	71.2	1.3	2478	3	US-08-485-355B-47 Sequence 47, Appl
7	71.2	1.3	2478	3	US-08-485-355B-49 Sequence 49, Appl
8	71.2	1.3	2478	3	US-09-194-613-1 Sequence 1, Appl
9	71.2	1.3	2479	3	US-08-485-355B-51 Sequence 51, Appl
10	50.8	1.0	1008	4	US-09-500-569-1 Sequence 1, Appl
11	47.6	0.9	1116	4	US-09-572-422A-41 Sequence 41, Appl
12	47.2	0.9	567	4	US-09-552-991A-3828 Sequence 3828, Ap
13	47.2	0.9	47981	4	US-09-679-279-1 Sequence 1, Appl
14	46.8	0.9	4403765	3	US-09-103-840A-2 Sequence 2, Appl
15	45.8	0.9	1193	4	US-09-372-422A-23 Sequence 23, Appl
16	45.6	0.9	447	4	US-09-252-991A-3732 Sequence 3732, Ap
17	45.4	0.9	1926	4	US-09-249-585A-4 Sequence 4, Appl
18	45.4	0.9	1931	2	US-09-130-114-2 Sequence 2, Appl
19	44.8	0.8	4411529	3	US-09-103-840A-1 Sequence 1, Appl
20	44.4	0.8	44377	2	US-08-804-227C-7 Sequence 7, Appl
21	44.4	0.8	77536	4	US-09-410-551B-1 Sequence 1, Appl
22	43.8	0.8	1304	4	US-09-372-422A-15 Sequence 15, Appl
23	43.4	0.8	1044	4	US-07-975-526-3 Sequence 3, Appl
24	43.4	0.8	1044	4	US-07-974-409C-425 Sequence 425, App
25	43.4	0.8	1935	4	US-09-620-312D-236 Sequence 236, App
26	43.4	0.8	1854	4	US-09-252-991A-14350 Sequence 14350, A
27	43.2	0.8	1854	4	US-09-252-991A-14350 Sequence 14350, A

28	43.2	0.8	2943	4	US-09-252-991A-14267 Sequence 14267, A
29	43.2	0.8	4257	2	US-08-690-473-1 Sequence 1, Appl
30	43.2	0.8	4257	3	US-09-259-821A-1 Sequence 1, Appl
31	43.2	0.8	4257	3	US-08-843-659-1 Sequence 1, Appl
32	42.6	0.8	4403765	3	US-09-103-840A-2 Sequence 2, Appl
33	42.6	0.8	4411529	3	US-09-103-840A-1 Sequence 1, Appl
34	42	0.8	1596	3	US-08-416-544B-8 Sequence 8, Appl
35	41.8	0.8	900	4	US-09-252-991A-5764 Sequence 5764, Ap
36	41.8	0.8	1575	3	US-08-478-507-11 Sequence 11, Appl
37	41.8	0.8	1575	3	US-09-128-275A-11 Sequence 11, Appl
38	41.8	0.8	1575	4	US-09-553-427-11 Sequence 11, Appl
39	41.8	0.8	1854	4	US-09-252-991A-5747 Sequence 5747, Ap
40	41.8	0.8	2085	4	US-09-252-991A-5708 Sequence 5708, Ap
41	41.8	0.8	7171	3	US-08-478-507-10 Sequence 10, Appl
42	41.8	0.8	7171	3	US-09-128-275A-10 Sequence 10, Appl
43	41.8	0.8	7171	4	US-09-553-427-10 Sequence 10, Appl
44	41.6	0.8	528	4	US-09-679-279-25 Sequence 25, Appl
45	41.6	0.8	1965	3	US-09-178-252-26 Sequence 26, Appl

ALIGNMENTS

RESULT 1
US-08-485-355B-39
Sequence 39, Application US/08485355B
Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Chirietian, P. D., Gordon, K. H.D., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESSES:
ADDRESSES: Flehr Hombach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 510 277299
INFORMATION FOR SEQ ID NO: 39:
SEQUENCE CHARACTERISTICS:
LENGTH: 5312 base pairs
TYPE: nucleic acid
STRANDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 37..5148

SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-08-485-355B-39

Query Match 100.0%; Score 5312; DB 3; Length 5312;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTTCTGCTCCCGGACGGTAAATATAGGGGAACTGTATGCGGAAAGCGACAGAGCTG 60
DB 1 GTTCTGCTCCCGGACGGTAAATATAGGGGAACTGTATGCGGAAAGCGACAGAGCTG 60
QY 61 GCGCGGTCTACCGCGGAGAGATGTCCTTACCGGAAAGCTACTGACGAGAGAGAGT 120
DB 61 GCGCGGTCTACCGCGGAGAGATGTCCTTACCGGAAAGCTACTGACGAGAGAGAGT 120
QY 121 AAGTGGACTTCCCGCGGACCTAGAGGCACTAGAAAACCTCCACAGACTGTACTACG 180
DB 121 AAGTGGACTTCCCGCGGACCTAGAGGCACTAGAAAACCTCCACAGACTGTACTACG 180
QY 181 CTGCGCTTCAAGGGGGCACTTTACCCCGACACAACACCCGATCTGCGCGGACCAA 240
DB 181 CTGCGCTTCAAGGGGGCACTTTACCCCGACACAACACCCGATCTGCGCGGACCAA 240
QY 241 CGGTGCGCAAAAGAGTTCTGCACAATTTGCGGAGGAGTGAAGCAAGTGTGAGATA 300
DB 241 CGGTGCGCAAAAGAGTTCTGCACAATTTGCGGAGGAGTGAAGCAAGTGTGAGATA 300
QY 301 GGGCGGCTCTGCAAGCGCACTTAAGCTACATGAGGACCGAAACGCGCGCTGCGAGAC 360
DB 301 GGGCGGCTCTGCAAGCGCACTTAAGCTACATGAGGACCGAAACGCGCGCTGCGAGAC 360
QY 361 TATCACGGGTGACCAAGTACGCGACCCGCGAGCTGCGGACACATTAACGCTTAAAG 420
DB 361 TATCACGGGTGACCAAGTACGCGACCCGCGAGCTGCGGACACATTAACGCTTAAAG 420
QY 421 TCTAGATCCGTCGCAAGCGCGCGCGGAGTTCAAGGCGGACGCTGCTGCGCAAC 480
DB 421 TCTAGATCCGTCGCAAGCGCGCGCGGAGTTCAAGGCGGACGCTGCTGCGCAAC 480
QY 481 GGCATTGCTCCGCGACCTTCTGCGTGAAGAGTGGGCTTTGCGCGGTTCAATCGCG 540
DB 481 GGCATTGCTCCGCGACCTTCTGCGTGAAGAGTGGGCTTTGCGCGGTTCAATCGCG 540
QY 541 GTTGAATTTGCCAATCACTCCCTTATGACGTGACCTTAAGAGAGTGGCCATGCGTT 600
DB 541 GTTGAATTTGCCAATCACTCCCTTATGACGTGACCTTAAGAGAGTGGCCATGCGTT 600
QY 601 GAGAACCAAGGACTTCAATGTCGCGCGGCTTCAATGCAATGCGAAGAGAGTGTCTAC 660
DB 601 GAGAACCAAGGACTTCAATGTCGCGCGGCTTCAATGCAATGCGAAGAGAGTGTCTAC 660
QY 661 ATGGAACAAGTGTAAATGCGAGCTCGCTACCGCTTCCAGTTATTGAAGAGCTATG 720
DB 661 ATGGAACAAGTGTAAATGCGAGCTCGCTACCGCTTCCAGTTATTGAAGAGCTATG 720
QY 721 GCTGTGAAGAGCTGCGATTCAGAGGAGGAGGAGCTTCCGTTCTCACTTCCGAGTTG 780
DB 721 GCTGTGAAGAGCTGCGATTCAGAGGAGGAGGAGCTTCCGTTCTCACTTCCGAGTTG 780
QY 781 TTTCTAACCAAGAGGAGGAGGAGGAGTGAAGAGGCTGCGCGCGGCTCTACTCC 840
DB 781 TTTCTAACCAAGAGGAGGAGGAGGAGTGAAGAGGCTGCGCGCGGCTCTACTCC 840
QY 841 AGAGCGCGCTCATTTTCTCGCGGAGAGAGAGTGGAGTGAAGTGAAGTGAAGTGA 900
DB 841 AGAGCGCGCTCATTTTCTCGCGGAGAGAGAGTGGAGTGAAGTGAAGTGAAGTGA 900
QY 901 CACACATGAGCTCGCTTACCTAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 960
DB 901 CACACATGAGCTCGCTTACCTAATGAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 960
QY 961 CATTAAGAGTCCAGAGGCGGACAGGCTCGAGATTAAGTGGGATCACTCGGCGGCA 1020
DB 961 CATTAAGAGTCCAGAGGCGGACAGGCTCGAGATTAAGTGGGATCACTCGGCGGCA 1020

```

```

DB 961 CATTAAGAGTCCAGAGGCGGACAGGCTCGAGATTAAGTGGGATCACTCGGCGGCA 1020
QY 1021 CCTGAGAGCGGATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
DB 1021 CCTGAGAGCGGATGCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080
QY 1081 AACATCTTTTATTAACCGGACCGCTGCGGACCTGAGCAATTAACCACTCTTACGTCAG 1140
DB 1081 AACATCTTTTATTAACCGGACCGCTGCGGACCTGAGCAATTAACCACTCTTACGTCAG 1140
QY 1141 CACAAAGTCAACATGCTGCTCAATTTATGCAAAACGCTCTGAGAGAGAACTAGTCAG 1200
DB 1141 CACAAAGTCAACATGCTGCTCAATTTATGCAAAACGCTCTGAGAGAGAACTAGTCAG 1200
QY 1201 ATGACCGTCTTGAATGCTTTCGCGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1260
DB 1201 ATGACCGTCTTGAATGCTTTCGCGCGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAG 1260
QY 1261 GTTACCGGAGAGTCTGAGAACTTCAACCGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1320
DB 1261 GTTACCGGAGAGTCTGAGAACTTCAACCGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1320
QY 1321 TACGTCCTCAATCATGAGGCGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1380
DB 1321 TACGTCCTCAATCATGAGGCGGAGGCTGAGGCTGAGGCTGAGGCTGAGGCTGAGGCT 1380
QY 1381 GACGTCCTTGAAGAGCTTCTGTTGAGAGAGTCTCAAGCAAGTCTTGAAGAGTCTT 1440
DB 1381 GACGTCCTTGAAGAGCTTCTGTTGAGAGAGTCTCAAGCAAGTCTTGAAGAGTCTT 1440
QY 1441 GGTCTGCGGACCTCAAGAGGCGGAGGCTGCTTCAAGAGGCGGCTGCTTCAAGAGG 1500
DB 1441 GGTCTGCGGACCTCAAGAGGCGGAGGCTGCTTCAAGAGGCGGCTGCTTCAAGAGG 1500
QY 1501 CGAGTCACTGCTGCGGAGCAATATCTGAGAGTCTGCGGCTGCTGCGGCTGCGGCTG 1560
DB 1501 CGAGTCACTGCTGCGGAGCAATATCTGAGAGTCTGCGGCTGCTGCGGCTGCGGCTG 1560
QY 1561 TTCCTGCGGCTGCGGAGCACTTCTGAGAGGCTGCTTCAAGAGGCGGCTGCTTCAAG 1620
DB 1561 TTCCTGCGGCTGCGGAGCACTTCTGAGAGGCTGCTTCAAGAGGCGGCTGCTTCAAG 1620
QY 1621 CTTCGGAAGCTGCGGCTGCTGAGAGGCTGCTTCAAGAGGCTGCTTCAAGAGGCTG 1680
DB 1621 CTTCGGAAGCTGCGGCTGCTGAGAGGCTGCTTCAAGAGGCTGCTTCAAGAGGCTG 1680
QY 1681 CCCCAAGTTCGAGCGGCACTTGTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 1740
DB 1681 CCCCAAGTTCGAGCGGCACTTGTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 1740
QY 1741 GCCACCTTGAAGGAGGCTGCAACCGGAGCTGAGAGGCTGAGAGGCTGAGAGGCT 1800
DB 1741 GCCACCTTGAAGGAGGCTGCAACCGGAGCTGAGAGGCTGAGAGGCTGAGAGGCT 1800
QY 1801 GGCCTCAAGGAGGCTGCAACCGGAGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAG 1860
DB 1801 GGCCTCAAGGAGGCTGCAACCGGAGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAG 1860
QY 1861 GGCCTCAAGGAGGCTGCAACCGGAGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAG 1920
DB 1861 GGCCTCAAGGAGGCTGCAACCGGAGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAG 1920
QY 1921 AAAGCACTTAACTGAGGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 1980
DB 1921 AAAGCACTTAACTGAGGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 1980
QY 1981 CCGCGCTGCGGCTGAGGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 2040
DB 1981 CCGCGCTGCGGCTGAGGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 2040
QY 2041 GAGGCGCGGCTTCTGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 2100
DB 2041 GAGGCGCGGCTTCTGCTGAGGCTGAGAGGCTGAGAGGCTGAGAGGCTGAGAGGCT 2100

```

QY 2101 GTGGGATCGTGCAGCGCTTGTGCCGAGCTCAGCAATAGTCTTGTAGGGAGCGTCCAC 2160
 DB 2101 GTGGGATCGTGCAGCGCTTGTGCCGAGCTCAGCAATAGTCTTGTAGGGAGCGTCCAC 2160
 QY 2161 CAATCGGGTTATATAGACTTCCAGAGCAGACGCGCAATGCGCTCGTTCGCGAGCTC 2220
 DB 2161 CAATCGGGTTATATAGACTTCCAGAGCAGACGCGCAATGCGCTCGTTCGCGAGCTC 2220
 QY 2221 GTTAGAGAGTGGCGGTGCGGCACTTTCACCAACCAAGCGCTGTCGGGCGAGCTCGTT 2280
 DB 2221 GTTAGAGAGTGGCGGTGCGGCACTTTCACCAACCAAGCGCTGTCGGGCGAGCTCGTT 2280
 QY 2281 GGCACACGTTTTTTCAGAGCTTGTACCCCGGGTGCACCAACCACTCAGGGTGCCTGCA 2340
 DB 2281 GGCACACGTTTTTTCAGAGCTTGTACCCCGGGTGCACCAACCACTCAGGGTGCCTGCA 2340
 QY 2341 TCCATCAGCAGTGCAGCGCCAGACTACCGCAACAGCGAGCGCAACCGCTGCTTCAAG 2400
 DB 2341 TCCATCAGCAGTGCAGCGCCAGACTACCGCAACAGCGAGCGCAACCGCTGCTTCAAG 2400
 QY 2401 CAGGAGGAAAAGTCCGCGCCACGCGGCTGAGGGCGGAGTACTGTGCAAGAGCGAGGGA 2460
 DB 2401 CAGGAGGAAAAGTCCGCGCCACGCGGCTGAGGGCGGAGTACTGTGCAAGAGCGAGGGA 2460
 QY 2461 CGCAGCTTTGCGTCTGTCACTTGCATTACCAAGCGCTCAGAGCAGAGGAGGCTCTC 2520
 DB 2461 CGCAGCTTTGCGTCTGTCACTTGCATTACCAAGCGCTCAGAGCAGAGGAGGCTCTC 2520
 QY 2521 GCTGAGAGTGCACCTTCTAGTCGAGCATCAGCGCCACACCAACCTGTACATCCGC 2580
 DB 2521 GCTGAGAGTGCACCTTCTAGTCGAGCATCAGCGCCACACCAACCTGTACATCCGC 2580
 QY 2581 GACCCGACAGGTGACATTGAGAGCACTCAACCATAGCGCAAGCGAGGTGTTTACA 2640
 DB 2581 GACCCGACAGGTGACATTGAGAGCACTCAACCATAGCGCAAGCGAGGTGTTTACA 2640
 QY 2641 GACATCCCTGCAAGCTGAGATCAGACTGTCAACCGAGTGAAGAGGTGAGCGCAAC 2700
 DB 2641 GACATCCCTGCAAGCTGAGATCAGACTGTCAACCGAGTGAAGAGGTGAGCGCAAC 2700
 QY 2701 GAAAGTATGCAACGATACCCCGGAGAGTGCACGCGCCACGAGCAATCCATCTGCTC 2760
 DB 2701 GAAAGTATGCAACGATACCCCGGAGAGTGCACGCGCCACGAGCAATCCATCTGCTC 2760
 QY 2761 CGCAAGAACTTGGGGACCAACCCGACTGTGCTGTGCTTTGGCGAAGACCGCTAC 2820
 DB 2761 CGCAAGAACTTGGGGACCAACCCGACTGTGCTGTGCTTTGGCGAAGACCGCTAC 2820
 QY 2821 GAGGTGTTGGGGGTCTGTCGCAAAATCAACGTAAGCTTGGCCGAACCGGACCGCCG 2880
 DB 2821 GAGGTGTTGGGGGTCTGTCGCAAAATCAACGTAAGCTTGGCCGAACCGGACCGCCG 2880
 QY 2881 AAGCCGATAGGGCGTTCAGAGAGGGGATCAGTGGGTCAAGTCCCAACCGCTTAC 2940
 DB 2881 AAGCCGATAGGGCGTTCAGAGAGGGGATCAGTGGGTCAAGTCCCAACCGCTTAC 2940
 QY 2941 AAAACACAGGCGCTCAGAGCTGTTGTCCTGCTACACCAAGCGAAGCGCTGCGG 3000
 DB 2941 AAAACACAGGCGCTCAGAGCTGTTGTCCTGCTACACCAAGCGAAGCGCTGCGG 3000
 QY 3001 CTACACGAACTAAGAGAGCTGCAAAACGATGCTAACTCGTTGACCGAGATTGGGAC 3060
 DB 3001 CTACACGAACTAAGAGAGCTGCAAAACGATGCTAACTCGTTGACCGAGATTGGGAC 3060
 QY 3061 TGAAGTGTCACTGAAGACCGCGGTGACGAGCTGTCTTGAAGACCGAGCTCAAGTTAC 3120
 DB 3061 TGAAGTGTCACTGAAGACCGCGGTGACGAGCTGTCTTGAAGACCGAGCTCAAGTTAC 3120
 QY 3121 CAAGCGGCGGACCGTGCAGAGACTGTGAGAGCAGACCGCTTCACTCCGTGACATA 3180
 DB 3121 CAAGCGGCGGACCGTGCAGAGACTGTGAGAGCAGACCGCTTCACTCCGTGACATA 3180
 QY 3181 GACTTCCTTATGAGAGCTCAGAGAAAGTGTGCGCCAGCGAGTCAATACGGGAGAGTC 3240
 DB 3181 GACTTCCTTATGAGAGCTCAGAGAAAGTGTGCGCCAGCGAGTCAATACGGGAGAGTC 3240
 QY 3241 GGGCAGGGAGATCGCGCTCACTCAAAAGTCTTCAACCTTGTCTCGCGCTTGAATACGC 3300
 DB 3241 GGGCAGGGAGATCGCGCTCACTCAAAAGTCTTCAACCTTGTCTCGCGCTTGAATACGC 3300
 QY 3301 ATACTCGAGAGATCTCTCGTACCGGGAGCGGACCGGCTCGGTACAGCAACGCTTCCC 3360
 DB 3301 ATACTCGAGAGATCTCTCGTACCGGGAGCGGACCGGCTCGGTACAGCAACGCTTCCC 3360
 QY 3361 GACGAGAGAGAGCCATGCTCTCGAGAGAGATCAATCAAGTCCCAACCGCAGCTTC 3420
 DB 3361 GACGAGAGAGAGCCATGCTCTCGAGAGAGATCAATCAAGTCCCAACCGCAGCTTC 3420
 QY 3421 GTTCGCGGAGCTGAGCCGAGTTGACACCGCCCAATTAACAGAGTGAAGTCTTTC 3480
 DB 3421 GTTCGCGGAGCTGAGCCGAGTTGACACCGCCCAATTAACAGAGTGAAGTCTTTC 3480
 QY 3481 GCGCGCTTTTAAAGCGGATGCGGACCGCTGCAAGTGCCTGTAATCTATTCAGAGAACG 3540
 DB 3481 GCGCGCTTTTAAAGCGGATGCGGACCGCTGCAAGTGCCTGTAATCTATTCAGAGAACG 3540
 QY 3541 TGTGGAAAACGCACTTGCAGAGCGAAGGAGTCTCGTTGAAGTGAAGTGAAGTCTGCTC 3600
 DB 3541 TGTGGAAAACGCACTTGCAGAGCGAAGGAGTCTCGTTGAAGTGAAGTGAAGTCTGCTC 3600
 QY 3601 GACTCGGCGGAGCTTGAAGCGCTTGCAGCAACACATCTTCTGCGCGCTCATGCTC 3660
 DB 3601 GACTCGGCGGAGCTTGAAGCGCTTGCAGCAACACATCTTCTGCGCGCTCATGCTC 3660
 QY 3661 AGCTCTTCCGCGGGGTCAAGTTCGCAAGCTTCAAAAGCGAGACTCGCTCTCTGTGGT 3720
 DB 3661 AGCTCTTCCGCGGGGTCAAGTTCGCAAGCTTCAAAAGCGAGACTCGCTCTCTGTGGT 3720
 QY 3721 AGCCATTACCTCCGTTGAGAGCTAGCGGCTTCAATGAGGCGAAGTTACAAAGCAAA 3780
 DB 3721 AGCCATTACCTCCGTTGAGAGCTAGCGGCTTCAATGAGGCGAAGTTACAAAGCAAA 3780
 QY 3781 CATTGAAAGTGTGAGTGCAGAAATGTCGCGTACATCGACTCTGCTCGCTGAG 3840
 DB 3781 CATTGAAAGTGTGAGTGCAGAAATGTCGCGTACATCGACTCTGCTCGCTGAG 3840
 QY 3841 CAGGTCTCTGCAACCTGTGAGAGCGCTCTCAAGATATTTGGCGCTGCTCAACAAC 3900
 DB 3841 CAGGTCTCTGCAACCTGTGAGAGCGCTCTCAAGATATTTGGCGCTGCTCAACAAC 3900
 QY 3901 GAACTCCTTACTCAAGTACGAGGCTGTGAGAGCATGCAAGAGGCTGAGGTAC 3960
 DB 3901 GAACTCCTTACTCAAGTACGAGGCTGTGAGAGCATGCAAGAGGCTGAGGTAC 3960
 QY 3961 GCCCGCTACACAGCTCTGTCGACATGTCAGATGCTACTCAATTAACGCGCGAG 4020
 DB 3961 GCCCGCTACACAGCTCTGTCGACATGTCAGATGCTACTCAATTAACGCGCGAG 4020
 QY 4021 TCTGCGGGATCATATGAGCGCTGTGTCTGCTTTGGCGGCGGAGACTTCCGTTGAA 4080
 DB 4021 TCTGCGGGATCATATGAGCGCTGTGTCTGCTTTGGCGGCGGAGACTTCCGTTGAA 4080
 QY 4081 CAATCGCGGGTGTGCGGCAATGAGAGGACCGGACGCTTACAGAGCAGTATCCG 4140
 DB 4081 CAATCGCGGGTGTGCGGCAATGAGAGGACCGGACGCTTACAGAGCAGTATCCG 4140
 QY 4141 GCTAACGTGCGGCACTGTGCTTGAACAGCTTTCGAGCCCGGACGCGCGCGCCG 4200
 DB 4141 GCTAACGTGCGGCACTGTGCTTGAACAGCTTTCGAGCCCGGACGCGCGCGCCG 4200
 QY 4201 GAGGTTTCTGTCGACATGTGCGAAGCGGAAAGCGCTTCTTCAACCGCAAGGT 4260
 DB 4201 GAGGTTTCTGTCGACATGTGCGAAGCGGAAAGCGCTTCTTCAACCGCAAGGT 4260
 QY 4261 GGTGTTCTGCACTAACAGCAAGTTCGAGCTGGGACTGGGCGCGCGAGTCTCATGG 4320

QY 3181 GACTTCCTTATGAGAGCTCAGAGAAAGTGTGCGCCAGCGAGTCAATACGGGAGAGTC 3240
 DB 3181 GACTTCCTTATGAGAGCTCAGAGAAAGTGTGCGCCAGCGAGTCAATACGGGAGAGTC 3240
 QY 3241 GGGCAGGGAGATCGCGCTCACTCAAAAGTCTTCAACCTTGTCTCGCGCTTGAATACGC 3300
 DB 3241 GGGCAGGGAGATCGCGCTCACTCAAAAGTCTTCAACCTTGTCTCGCGCTTGAATACGC 3300
 QY 3301 ATACTCGAGAGATCTCTCGTACCGGGAGCGGACCGGCTCGGTACAGCAACGCTTCCC 3360
 DB 3301 ATACTCGAGAGATCTCTCGTACCGGGAGCGGACCGGCTCGGTACAGCAACGCTTCCC 3360
 QY 3361 GACGAGAGAGAGCCATGCTCTCGAGAGAGATCAATCAAGTCCCAACCGCAGCTTC 3420
 DB 3361 GACGAGAGAGAGCCATGCTCTCGAGAGAGATCAATCAAGTCCCAACCGCAGCTTC 3420
 QY 3421 GTTCGCGGAGCTGAGCCGAGTTGACACCGCCCAATTAACAGAGTGAAGTCTTTC 3480
 DB 3421 GTTCGCGGAGCTGAGCCGAGTTGACACCGCCCAATTAACAGAGTGAAGTCTTTC 3480
 QY 3481 GCGCGCTTTTAAAGCGGATGCGGACCGCTGCAAGTGCCTGTAATCTATTCAGAGAACG 3540
 DB 3481 GCGCGCTTTTAAAGCGGATGCGGACCGCTGCAAGTGCCTGTAATCTATTCAGAGAACG 3540
 QY 3541 TGTGGAAAACGCACTTGCAGAGCGAAGGAGTCTCGTTGAAGTGAAGTGAAGTCTGCTC 3600
 DB 3541 TGTGGAAAACGCACTTGCAGAGCGAAGGAGTCTCGTTGAAGTGAAGTGAAGTCTGCTC 3600
 QY 3601 GACTCGGCGGAGCTTGAAGCGCTTGCAGCAACACATCTTCTGCGCGCTCATGCTC 3660
 DB 3601 GACTCGGCGGAGCTTGAAGCGCTTGCAGCAACACATCTTCTGCGCGCTCATGCTC 3660
 QY 3661 AGCTCTTCCGCGGGGTCAAGTTCGCAAGCTTCAAAAGCGAGACTCGCTCTCTGTGGT 3720
 DB 3661 AGCTCTTCCGCGGGGTCAAGTTCGCAAGCTTCAAAAGCGAGACTCGCTCTCTGTGGT 3720
 QY 3721 AGCCATTACCTCCGTTGAGAGCTAGCGGCTTCAATGAGGCGAAGTTACAAAGCAAA 3780
 DB 3721 AGCCATTACCTCCGTTGAGAGCTAGCGGCTTCAATGAGGCGAAGTTACAAAGCAAA 3780
 QY 3781 CATTGAAAGTGTGAGTGCAGAAATGTCGCGTACATCGACTCTGCTCGCTGAG 3840
 DB 3781 CATTGAAAGTGTGAGTGCAGAAATGTCGCGTACATCGACTCTGCTCGCTGAG 3840
 QY 3841 CAGGTCTCTGCAACCTGTGAGAGCGCTCTCAAGATATTTGGCGCTGCTCAACAAC 3900
 DB 3841 CAGGTCTCTGCAACCTGTGAGAGCGCTCTCAAGATATTTGGCGCTGCTCAACAAC 3900
 QY 3901 GAACTCCTTACTCAAGTACGAGGCTGTGAGAGCATGCAAGAGGCTGAGGTAC 3960
 DB 3901 GAACTCCTTACTCAAGTACGAGGCTGTGAGAGCATGCAAGAGGCTGAGGTAC 3960
 QY 3961 GCCCGCTACACAGCTCTGTCGACATGTCAGATGCTACTCAATTAACGCGCGAG 4020
 DB 3961 GCCCGCTACACAGCTCTGTCGACATGTCAGATGCTACTCAATTAACGCGCGAG 4020
 QY 4021 TCTGCGGGATCATATGAGCGCTGTGTCTGCTTTGGCGGCGGAGACTTCCGTTGAA 4080
 DB 4021 TCTGCGGGATCATATGAGCGCTGTGTCTGCTTTGGCGGCGGAGACTTCCGTTGAA 4080
 QY 4081 CAATCGCGGGTGTGCGGCAATGAGAGGACCGGACGCTTACAGAGCAGTATCCG 4140
 DB 4081 CAATCGCGGGTGTGCGGCAATGAGAGGACCGGACGCTTACAGAGCAGTATCCG 4140
 QY 4141 GCTAACGTGCGGCACTGTGCTTGAACAGCTTTCGAGCCCGGACGCGCGCGCCG 4200
 DB 4141 GCTAACGTGCGGCACTGTGCTTGAACAGCTTTCGAGCCCGGACGCGCGCGCCG 4200
 QY 4201 GAGGTTTCTGTCGACATGTGCGAAGCGGAAAGCGCTTCTTCAACCGCAAGGT 4260
 DB 4201 GAGGTTTCTGTCGACATGTGCGAAGCGGAAAGCGCTTCTTCAACCGCAAGGT 4260
 QY 4261 GGTGTTCTGCACTAACAGCAAGTTCGAGCTGGGACTGGGCGCGCGAGTCTCATGG 4320

```

Db      4261 GGGTGTTCGCGACTCAAGCCAGTGGCACTGGAGCTCGCCCGGAGTCTCCATGG 4320
Qy      4321 GATGCACTGCGACCAACAGCTTTTGGAGTATTGACACCGGAGACCCCGTCACATCA 4380
Db      4321 GATGCACTGCGACCAACAGCTTTTGGAGTATTGACACCGGAGACCCCGTCACATCA 4380
Qy      4381 TCGTCGCGCTCATGCTTTCATCGGAGCTCTTACATGCTGGAAGTCCCTCATGCT 4440
Db      4381 TCGTCGCGCTCATGCTTTCATCGGAGCTCTTACATGCTGGAAGTCCCTCATGCT 4440
Qy      4441 GGAGACACCGGACCAAGACCAAGAAAGTAAACAGGAGAAAGCCCTTGGCAAGACAG 4500
Db      4441 GGAGACACCGGACCAAGACCAAGAAAGTAAACAGGAGAAAGCCCTTGGCAAGACAG 4500
Qy      4501 CAATCAAGCTGCTGTAATGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560
Db      4501 CAATCAAGCTGCTGTAATGCTGGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 4560
Qy      4561 GCGCCCACTGCTCGAGCCCTCTCATTTTGGAAAAGAGTGGAGTGGAGCCGGG 4620
Db      4561 GCGCCCACTGCTCGAGCCCTCTCATTTTGGAAAAGAGTGGAGTGGAGCCGGG 4620
Qy      4621 CCGACTGCTCGCGTGGAGATCACTTGGGCAACCCCATCTGCGGCAAGAGAGT 4680
Db      4621 CCGACTGCTCGCGTGGAGATCACTTGGGCAACCCCATCTGCGGCAAGAGAGT 4680
Qy      4681 GCGCGAGAGACAGCGCGGAGCTTTGCGGAGTCCAGTCTGCGCGGAGTCTGTC 4740
Db      4681 GCGCGAGAGACAGCGCGGAGCTTTGCGGAGTCCAGTCTGCGCGGAGTCTGTC 4740
Qy      4741 TCGAAGCGGAGCGGCGCCCAAGAGTCCGCTTGGAAAAGAGTCACTCCGCTCAGCG 4800
Db      4741 TCGAAGCGGAGCGGCGCCCAAGAGTCCGCTTGGAAAAGAGTCACTCCGCTCAGCG 4800
Qy      4801 TGGAAAGGCGTGAACCGGAGACCGGAGAGTCAAGGAGAGAGAGAGAGAGAGAG 4860
Db      4801 TGGAAAGGCGTGAACCGGAGACCGGAGAGTCAAGGAGAGAGAGAGAGAGAGAG 4860
Qy      4861 GTCCAGGCTCTGATCAGCGCGGCTTATCTCAGAAAGAGAGAGTCTTCTCCAGCATCC 4920
Db      4861 GTCCAGGCTCTGATCAGCGCGGCTTATCTCAGAAAGAGAGAGTCTTCTCCAGCATCC 4920
Qy      4921 AAAGGCTACTCAAGAACTAAAGGATGCTCAATCACTCTTCTGCGCCGAGTGGC 4980
Db      4921 AAAGGCTACTCAAGAACTAAAGGATGCTCAATCACTCTTCTGCGCCGAGTGGC 4980
Qy      4981 GATTACCAAGCGCGCGAGCTGCGAGACAGTCCGAGTCTGCGCGCGCTGCGAGATGGC 5040
Db      4981 GATTACCAAGCGCGCGAGCTGCGAGACAGTCCGAGTCTGCGCGCGCTGCGAGATGGC 5040
Qy      5041 CGCTCATGTATTCAGAGCGGCTTCACTGCGCGAGTGGCGAGCTTGAAGCGGATA 5100
Db      5041 CGCTCATGTATTCAGAGCGGCTTCACTGCGCGAGTGGCGAGCTTGAAGCGGATA 5100
Qy      5101 CGCTCATGTATTCAGAGCGGCTTCACTGCGCGAGTGGCGAGCTTGAAGCGGATA 5160
Db      5101 CGCTCATGTATTCAGAGCGGCTTCACTGCGCGAGTGGCGAGCTTGAAGCGGATA 5160
Qy      5161 ATTAGTTCTCTTGTGTAACCAAGGTGCTCCCTCCATGAGGTAAAGACTCTGCTGAG 5220
Db      5161 ATTAGTTCTCTTGTGTAACCAAGGTGCTCCCTCCATGAGGTAAAGACTCTGCTGAG 5220
Qy      5221 TCTCAAGCTTACTGTTGAGTCTGCTGCGGTTGATTCATTCCTCAAGAGCAAGAGGT 5280
Db      5221 TCTCAAGCTTACTGTTGAGTCTGCTGCGGTTGATTCATTCCTCAAGAGCAAGAGGT 5280
Qy      5281 GCGCACTAGTAGAGGCGCGCCCTGAGGATACCA 5312
Db      5281 GCGCACTAGTAGAGGCGCGCCCTGAGGATACCA 5312

```

RESULT 2

```

US-08-485-355B-41
; Sequence 41, Application US/08485355B
; Patent No. 6177075
; GENERAL INFORMATION:
; APPLICANT: Christlan, P. D., Gordon, K. H.J., Hanzlik, T. N.
; TITLE OF INVENTION: Insect Viruses and Their Uses in
; Protecting Plants
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Flehr Hobbach Test Albriltton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/485,355B
; FILING DATE: 07-Jun-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/440,522
; FILING DATE: 12-MAY-1995
; APPLICATION NUMBER: US 08/089,372
; FILING DATE: 08-JUL-1993
; APPLICATION NUMBER: AU PL4081/92
; FILING DATE: 14-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 761-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5312 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4218..4514
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-08-485-355B-41
Query Match 100.0%; Score 5312; DB 3; Length 5312;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy      1 GTTCTGCTCCCGGACGGTAAATATAGGGGAAACATGTACCGGAAGGAGACAGAGTGG 60
Db      1 GTTCTGCTCCCGGACGGTAAATATAGGGGAAACATGTACCGGAAGGAGACAGAGTGG 60
Qy      61 GCGCGTGTCTAAGCGCGGAGAGTGTGCTTACGCGGAAGCTACTGCGAGAGAGAGTGC 120
Db      61 GCGCGTGTCTAAGCGCGGAGAGTGTGCTTACGCGGAAGCTACTGCGAGAGAGAGTGC 120
Qy      121 AAGTTGACTTGGCGCGGACCTGAAGGACCTGAAGAACTCTCCACAGACTGTACTATCCG 180
Db      121 AAGTTGACTTGGCGCGGACCTGAAGGACCTGAAGAACTCTCCACAGACTGTACTATCCG 180
Qy      181 CTGCGCTTCAAGAGGAGGACCTTACCCCGGACACCAACCCGATCTCTGCGCGGACCAA 240
Db      181 CTGCGCTTCAAGAGGAGGACCTTACCCCGGACACCAACCCGATCTCTGCGCGGACCAA 240
Qy      241 CGTGTGCAAGAAAGTCTTGCACAAATTTGCGCAAGGGAGAGTACAGACAGTGTGAGATA 300

```

Db 241 CGGTGCGAAGAGGTTCTGCAATTTGCGCAAGGGAGGTAGCAGTGTCTGAGATG 300
 QY 301 GGGCCGCTCTCTGCAAGCGCATTTAAAGCTACATGGGGCAAGAACGCCCCGTGCGAGAC 360
 Db 301 GGGCCGCTCTCTGCAAGCGCATTTAAAGCTACATGGGGCAAGAACGCCCCGTGCGAGAC 360
 QY 361 TATCACGGGTGACCAAGTACGGCACCGCGAGCGCTCGGACACATTAAGGCTTAAGAG 420
 Db 361 TATCACGGGTGACCAAGTACGGCACCGCGAGCGCTCGGACACATTAAGGCTTAAGAG 420
 QY 421 TCTAGATCCGTCGCGCACAGGGCCGCGAGTTCAAGGCGCAGCTCACTGCTCGCAAC 480
 Db 421 TCTAGATCCGTCGCGCACAGGGCCGCGAGTTCAAGGCGCAGCTCACTGCTCGCAAC 480
 QY 481 GGCATTGCTCCCGCACCTTCTGCGTGAAGAGTGGCTTTTGCGGCTTCAAAATCGCGC 540
 Db 481 GGCATTGCTCCCGCACCTTCTGCGTGAAGAGTGGCTTTTGCGGCTTCAAAATCGCGC 540
 QY 541 GTTGAATTGGCCAAATGACTCCCTCTATGACGTGACCCCTAGAGAGAGTGGCCAAATGCGTTT 600
 Db 541 GTTGAATTGGCCAAATGACTCCCTCTATGACGTGACCCCTAGAGAGAGTGGCCAAATGCGTTT 600
 QY 601 GAGAACCAAGGACTTCAATGATCGCGGCTTCAATGACATGCGAGAGAGTGTCTTAC 660
 Db 601 GAGAACCAAGGACTTCAATGATCGCGGCTTCAATGACATGCGAGAGAGTGTCTTAC 660
 QY 661 ATGGAACACGTGGTTAATGATCGCGGCTTCAATGACATGCGAGAGAGTGTCTTAC 720
 Db 661 ATGGAACACGTGGTTAATGATCGCGGCTTCAATGACATGCGAGAGAGTGTCTTAC 720
 QY 721 GCTGTGAAGAGACTGCGCATTTCCAGGGGGGGGAGCTCCGTCTCACTTCCCTGAGTTGAGAC 780
 Db 721 GCTGTGAAGAGACTGCGCATTTCCAGGGGGGGGAGCTCCGTCTCACTTCCCTGAGTTGAGAC 780
 QY 781 TTTCATCAACGAGAGCGCAAGAGCGGCGCATGAGAGGCTGGCGCGCGGCTCTTAC 840
 Db 781 TTTCATCAACGAGAGCGCAAGAGCGGCGCATGAGAGGCTGGCGCGCGGCTCTTAC 840
 QY 841 AGACGGCGCGCTATTTCTCCGGCGAGCGAGCATGCGGGGTGATGCGTACTTACACGACTTC 900
 Db 841 AGACGGCGCGCTATTTCTCCGGCGAGCGAGCATGCGGGGTGATGCGTACTTACACGACTTC 900
 QY 901 CACACATGCTGCTGCTACTGCTGAGAGAACTACCCCATCTCGTTGGTTTCTCACTC 960
 Db 901 CACACATGCTGCTGCTACTGCTGAGAGAACTACCCCATCTCGTTGGTTTCTCACTC 960
 QY 961 CATATAGAAATCAGAGGCGCGCACGCGCTCAGCATTTGAGCTGCGCATCACTGCGCGGCA 1020
 Db 961 CATATAGAAATCAGAGGCGCGCACGCGCTCAGCATTTGAGCTGCGCATCACTGCGCGGCA 1020
 QY 1021 CCTGAGAGCCGATGCTGCGCGCTGCTCCAGAGACGCTCCCAAGGCTCTGCGAAGATCCCA 1080
 Db 1021 CCTGAGAGCCGATGCTGCGCGCTGCTCCAGAGACGCTCCCAAGGCTCTGCGAAGATCCCA 1080
 QY 1081 AACATCTTTTATTAAGCGCGCATGCGGCTGAGCATTAAGACCATCTTACGTACAG 1140
 Db 1081 AACATCTTTTATTAAGCGCGCATGCGGCTGAGCATTAAGACCATCTTACGTACAG 1140
 QY 1141 CACAAAGTCAACATGCTGCTCAATTTTATGCAAAAGCGCTCGAGAAAGGAATAGTCGAC 1200
 Db 1141 CACAAAGTCAACATGCTGCTCAATTTTATGCAAAAGCGCTCGAGAAAGGAATAGTCGAC 1200
 QY 1201 ATGACCGCTCTGATGCTGTCGCGCGCTAGGCTGCGCGCATCGTGTGCTCGACAGAA 1260
 Db 1201 ATGACCGCTCTGATGCTGTCGCGCGCTAGGCTGCGCGCATCGTGTGCTCGACAGAA 1260
 QY 1261 GTTACCGGAGAGCTCTGGAACATCTCAACGGGCTGACCTGTGCGCATCTGTGTCTT 1320
 Db 1261 GTTACCGGAGAGCTCTGGAACATCTCAACGGGCTGACCTGTGCGCATCTGTGTCTT 1320
 QY 1321 TAGCTCTTCACATCATGAGGCGCGAAGGGCTGCGGCTGTGTCAAGACCGCCAAAGGAC 1380
 Db 1321 TAGCTCTTCACATCATGAGGCGCGAAGGGCTGCGGCTGTGTCAAGACCGCCAAAGGAC 1380

Db 1381 TACGTCTTCACATCATGAGGCGCGAAGGGCTGCGGCTGTGTCAAGACCGCCAAAGGAC 1380
 QY 1381 GACGTCTTTGAGAGACTTGTGTTCTGGAGAGTCTCAAGACGCTTGTGGGCTCTGTTCG 1440
 Db 1381 GACGTCTTTGAGAGACTTGTGTTCTGGAGAGTCTCAAGACGCTTGTGGGCTCTGTTCG 1440
 QY 1441 GGTCTGCGGAACCTCAAAAGGACCCGACGTGCTTTTACTTAACGCGGTCTGTCAATAGTAC 1500
 Db 1441 GGTCTGCGGAACCTCAAAAGGACCCGACGTGCTTTTACTTAACGCGGTCTGTCAATAGTAC 1500
 QY 1501 CGAGTCACTCGCTCGAGAGCATTAATCTGCGAGCTGCGCTGTCCCTGAAAGAGTGGAG 1560
 Db 1501 CGAGTCACTCGCTCGAGAGCATTAATCTGCGAGCTGCGCTGTCCCTGAAAGAGTGGAG 1560
 QY 1561 TTCCTGCGCTCCGCGTACCATGCTGCGGCTCTTCAAGACAGGGAAGAGTGAAGTTC 1620
 Db 1561 TTCCTGCGCTCCGCGTACCATGCTGCGGCTCTTCAAGACAGGGAAGAGTGAAGTTC 1620
 QY 1621 CTTGCGGAAGCTGGCTGTGCTCAAGAGAGTCCGCTTCACTCTCTCTGTGAGAGAG 1680
 Db 1621 CTTGCGGAAGCTGGCTGTGCTCAAGAGAGTCCGCTTCACTCTCTCTGTGAGAGAG 1680
 QY 1681 CCCCAAGGTTTCAAGCGCGACTTGTGCGAGCGGACCGGCGCTCACTCCCGAGTACGCGC 1740
 Db 1681 CCCCAAGGTTTCAAGCGCGACTTGTGCGAGCGGACCGGCGCTCACTCCCGAGTACGCGC 1740
 QY 1741 GGCACCTTTCAGAGAGGCTCAACACGACGTCAAGAGAGCTCAAGATCAACCTTCAGAAC 1800
 Db 1741 GGCACCTTTCAGAGAGGCTCAACACGACGTCAAGAGAGCTCAAGATCAACCTTCAGAAC 1800
 QY 1801 GGCCTTCAAGACATGAGCGGGCTCAACCTCTCCCGAGTCAAGAGGCTCGAATGTACAG 1860
 Db 1801 GGCCTTCAAGACATGAGCGGGCTCAACCTCTCCCGAGTCAAGAGGCTCGAATGTACAG 1860
 QY 1861 GGCCTTCAAGACATGAGCGGGCTCAACCTCTCCCGAGTCAAGAGGCTCGAATGTACAG 1860
 Db 1861 GGCCTTCAAGACATGAGCGGGCTCAACCTCTCCCGAGTCAAGAGGCTCGAATGTACAG 1860
 QY 1921 AAAGCATTTTACGTGCGACCCACAGAGAACTGAGAGAGCTATGAGACCGGCGGATCAAA 1980
 Db 1921 AAAGCATTTTACGTGCGACCCACAGAGAACTGAGAGAGCTATGAGACCGGCGGATCAAA 1980
 QY 1981 CCGCGCTCGGCTCGGCTACGGAACATGTCGCTTGCAGATTCTCGCTGTCACCGCC 2040
 Db 1981 CCGCGCTCGGCTCGGCTACGGAACATGTCGCTTGCAGATTCTCGCTGTCACCGCC 2040
 QY 2041 GAGGGCGCGCTTGTGCTACCGTGGTTTCAAGAGTCTTCAAGTCCCGCTCGGTAC 2100
 Db 2041 GAGGGCGCGCTTGTGCTACCGTGGTTTCAAGAGTCTTCAAGTCCCGCTCGGTAC 2100
 QY 2101 GTGCGATCTGACAGCGCTTGTCCCGAGCTCAAGAAATGCTCTGTAGGGAGAGTGCAC 2160
 Db 2101 GTGCGATCTGACAGCGCTTGTCCCGAGCTCAAGAAATGCTCTGTAGGGAGAGTGCAC 2160
 QY 2161 CAATTCGGGTTTATAGACTTCCAAAGGCAAGAGCGGAACATGCTCGTGTGCGACGTC 2220
 Db 2161 CAATTCGGGTTTATAGACTTCCAAAGGCAAGAGCGGAACATGCTCGTGTGCGACGTC 2220
 QY 2221 GTTAAAGAGTGCCTGCGCGACTTTCACCAAGAGCGTGTCCGGCGGACGTCGTT 2280
 Db 2221 GTTAAAGAGTGCCTGCGCGACTTTCACCAAGAGCGTGTCCGGCGGACGTCGTT 2280
 QY 2281 GGCACACGTTTTTTCAGAGCTTGTACCCCGGAGTCAACAACACTCAAGGGTGCCTGCA 2340
 Db 2281 GGCACACGTTTTTTCAGAGCTTGTGTACCCCGGAGTCAACAACACTCAAGGGTGCCTGCA 2340
 QY 2341 TCCATCAGCAGCTGTCGCGCGACGACCTACCGCAACAGCCAGGCGCAACCTCTGCTTCAAG 2400
 Db 2341 TCCATCAGCAGCTGTCGCGCGACGACCTACCGCAACAGCCAGGCGCAACCTCTGCTTCAAG 2400
 QY 2401 CAGAGAGAAAGTGCAGCGCGCTGAGGGCGGAGTGAAGTGTGCAAGAGCGGAGGGA 2460
 Db 2401 CAGAGAGAAAGTGCAGCGCGCTGAGGGCGGAGTGAAGTGTGCAAGAGCGGAGGGA 2460

QY 2461 CGCATTTCGTCGTCTGTCATTTGTCATTAACAAGGCTCCACAGACAGACGAAAGCTCTC 2520
 Db 2461 CGCATTTCGTCGTCTGTCATTTGTCATTAACAAGGCTCCACAGACAGACGAAAGCTCTC 2520
 QY 2521 GCTGAGAAAGTCGACCTTCTAGTCGAGATCAAGCGGCAACAACCCCTGTATCATCGC 2580
 Db 2521 GCTGAGAAAGTCGACCTTCTAGTCGAGATCAAGCGGCAACAACCCCTGTATCATCGC 2580
 QY 2581 GACCCGACAGGTGACATTTGAGAGACAACCTAACATAGCCGGAAGCGAGTGTTTACA 2640
 Db 2581 GACCCGACAGGTGACATTTGAGAGACAACCTAACATAGCCGGAAGCGAGTGTTTACA 2640
 QY 2641 GACATCCCTGCAACCCCTGAGATCAAGCTGTCAAACCGAGTGAAGGTGACGCGAAC 2700
 Db 2641 GACATCCCTGCAACCCCTGAGATCAAGCTGTCAAACCGAGTGAAGGTGACGCGAAC 2700
 QY 2701 GAAGTGTGCGACAGATACCCCGCAGAGTGCACGCGCGACGAGCAATTCATCTGCTC 2760
 Db 2701 GAAGTGTGCGACAGATACCCCGCAGAGTGCACGCGCGACGAGCAATTCATCTGCTC 2760
 QY 2761 CGCAGAACTTCGCGGAGCAACCCGACTGTGTGCTTGGCGAAGACCGGCTAC 2820
 Db 2761 CGCAGAACTTCGCGGAGCAACCCGACTGTGTGCTTGGCGAAGACCGGCTAC 2820
 QY 2821 GAGGTGTTGGCGGTGCTGCCAAATCAACGTAGAGCTTGGCGAACCCGACGCGAC 2880
 Db 2821 GAGGTGTTGGCGGTGCTGCCAAATCAACGTAGAGCTTGGCGAACCCGACGCGAC 2880
 QY 2881 AAGCGGATAGGGGCTTCAGGAAAGGGGTACAGTGGGTCAAGGTCAACAAGCGCTTAC 2940
 Db 2881 AAGCGGATAGGGGCTTCAGGAAAGGGGTACAGTGGGTCAAGGTCAACAAGCGCTTAC 2940
 QY 2941 AAACACAGGCGCTCCAGACGCTGTTCGCGCTACACAAGCAAGCGCTGACCTGCG 3000
 Db 2941 AAACACAGGCGCTCCAGACGCTGTTCGCGCTACACAAGCAAGCGCTGACCTGCG 3000
 QY 3001 CTACACAGAGCTTAAGAGAGAGCTCAAAAGCATGCTAACTGCTTGAACCGCATTTGGGAC 3060
 Db 3001 CTACACAGAGCTTAAGAGAGAGCTCAAAAGCATGCTAACTGCTTGAACCGCATTTGGGAC 3060
 QY 3061 TGGACTGTCACTGAAGAGCGCGGTGACCGAGCTGTTCGAGACCCAGCTAAAGTTAC 3120
 Db 3061 TGGACTGTCACTGAAGAGCGCGGTGACCGAGCTGTTCGAGACCCAGCTAAAGTTAC 3120
 QY 3121 CAAGCGGCGGCAACGCTGAGAGACCTGTGAGGCGACAGACCCCTACATCCGTGACATA 3180
 Db 3121 CAAGCGGCGGCAACGCTGAGAGACCTGTGAGGCGACAGACCCCTACATCCGTGACATA 3180
 QY 3181 GACTTCTTATGAAGACTCAGACGAAAGTGTGCGCCAAAGCGGATCAATACCGGAGATC 3240
 Db 3181 GACTTCTTATGAAGACTCAGACGAAAGTGTGCGCCAAAGCGGATCAATACCGGAGATC 3240
 QY 3241 GGGGAGGGGATCGCGCTCACTCAAAAGTCTCAACTTGTGCTGCGGCTTGGATAGC 3300
 Db 3241 GGGGAGGGGATCGCGCTCACTCAAAAGTCTCAACTTGTGCTGCGGCTTGGATAGC 3300
 QY 3301 ATACTCGAGAGATCTCCGTACCGGAGCGGCAAGCTCCGTTACAGCAAGCTCTCCC 3360
 Db 3301 ATACTCGAGAGATCTCCGTACCGGAGCGGCAAGCTCCGTTACAGCAAGCTCTCCC 3360
 QY 3361 GACGAAGAGAGGCTCATGCTCTCGAAGCGAAGATCAATCAAGTCCCAACGCGCACTTC 3420
 Db 3361 GACGAAGAGAGGCTCATGCTCTCGAAGCGAAGATCAATCAAGTCCCAACGCGCACTTC 3420
 QY 3421 GTCTCGGCGGATCGACCGATTTGACACCGCCCAATTAACGAGAGAGTGTCTTC 3480
 Db 3421 GTCTCGGCGGATCGACCGATTTGACACCGCCCAATTAACGAGAGAGTGTCTTC 3480
 QY 3481 GCCGCCCTTTTGAAGCGATCGGCAAGCTGAGCTGCGTTAATCTATTGAGAAACG 3540
 Db 3481 GCCGCCCTTTTGAAGCGATCGGCAAGCTGAGCTGCGTTAATCTATTGAGAAACG 3540

QY 3541 TGTGGGAAACGACCTTGGAGAGGAGGTCTAGGCTCCGTTGAAGTCAGCGGTCTGCTC 3600
 Db 3541 TGTGGGAAACGACCTTGGAGAGGAGGTCTAGGCTCCGTTGAAGTCAGCGGTCTGCTC 3600
 QY 3601 GACTCGGCGACAGCTTGAAGCGCTTGCAGCAACCATCTTCTGCGCGGTCAATGCTC 3660
 Db 3601 GACTCGGCGACAGCTTGAAGCGCTTGCAGCAACCATCTTCTGCGCGGTCAATGCTC 3660
 QY 3661 AGCTCTTCCGCGGCGTCAAGTTGCGAGCTTTCAAAAGGCGACAGCTGCTCTGTGTGT 3720
 Db 3661 AGCTCTTCCGCGGCGTCAAGTTGCGAGCTTTCAAAAGGCGACAGCTGCTCTGTGTGT 3720
 QY 3721 AGCATTAACCTCGGTTTGAAGCTAGCGGCTTCAATGCGGAGGTTTCAAGAACCA 3780
 Db 3721 AGCATTAACCTCGGTTTGAAGCTAGCGGCTTCAATGCGGAGGTTTCAAGAACCA 3780
 QY 3781 CATTTGAAGGTGAGAGTGAAGAAATCTGTCGCTACATCGGACTCTCTGCTGAG 3840
 Db 3781 CATTTGAAGGTGAGAGTGAAGAAATCTGTCGCTACATCGGACTCTCTGCTGAG 3840
 QY 3841 GAGTGTCTCGACCTGTGAGAGCGCTTCAAGATATTTGGGCGTCTACACAAGC 3900
 Db 3841 GAGTGTCTCGACCTGTGAGAGCGCTTCAAGATATTTGGGCGTCTACACAAGC 3900
 QY 3901 GAACTCCTTACTCCAAATACGTAGAGGCTGTGAGAGCATCAACAAAGGCTGGAGTAC 3960
 Db 3901 GAACTCCTTACTCCAAATACGTAGAGGCTGTGAGAGCATCAACAAAGGCTGGAGTAC 3960
 QY 3961 GCCCGTACCAAGCTCTGTGCAATGTGACATGTGACATGCTACTTACATTAAGCGCGAG 4020
 Db 3961 GCCCGTACCAAGCTCTGTGCAATGTGACATGTGACATGCTACTTACATTAAGCGCGAG 4020
 QY 4021 TCTGCGGCGTACATCATGACGCTGTTGTGCTTGGCGCGGCGACTTCCGTTGAA 4080
 Db 4021 TCTGCGGCGTACATCATGACGCTGTTGTGCTTGGCGCGGCGACTTCCGTTGAA 4080
 QY 4081 CAACGCGGCGTGTGCGTGCACATGTGACAGGACCCGACTTACACAGACAGTATCCG 4140
 Db 4081 CAACGCGGCGTGTGCGTGCACATGTGACAGGACCCGACTTACACAGACAGTATCCG 4140
 QY 4141 GCTAACGTCGCGCATGTCCTTGAACAAGCTTTCAGAGCCCGCAGGCGCGCCCG 4200
 Db 4141 GCTAACGTCGCGCATGTCCTTGAACAAGCTTTCAGAGCCCGCAGGCGCGCCCG 4200
 QY 4201 GCAGCTTTCGTTGCAATGTGAGAGCGGAAAAGCTTCTTCACTTACCGGAAAGCT 4260
 Db 4201 GCAGCTTTCGTTGCAATGTGAGAGCGGAAAAGCTTCTTCACTTACCGGAAAGCT 4260
 QY 4261 GGTGTTCTGCGACTAACAGCAAGTGTGCACTGGGACTGGCGCCGGAAGTCTCAATGG 4320
 Db 4261 GGTGTTCTGCGACTAACAGCAAGTGTGCACTGGGACTGGCGCCGGAAGTCTCAATGG 4320
 QY 4321 GATGCACTGAGCAACAGCTTTTGGAGTTATTGACCGGAGACCCGCTCCACATCA 4380
 Db 4321 GATGCACTGAGCAACAGCTTTTGGAGTTATTGACCGGAGACCCGCTCCACATCA 4380
 QY 4381 TCTTCGCGGTATGCTTCTTATCGGACTCTCTTACATGTGTGAAAGGTCTGATGTGT 4440
 Db 4381 TCTTCGCGGTATGCTTCTTATCGGACTCTCTTACATGTGTGAAAGGTCTGATGTGT 4440
 QY 4441 GGAAGACCCGGAAGACACAGAGACTTGAACAGGAGAAAGCGGCTTGGCAAGACAG 4500
 Db 4441 GGAAGACCCGGAAGACACAGAGACTTGAACAGGAGAAAGCGGCTTGGCAAGACAG 4500
 QY 4501 CAATCAAGCTGTCTGAATGTCTGAGACAGAGCGGAGAAAGACAGGAGTTCGTTACT 4560
 Db 4501 CAATCAAGCTGTCTGAATGTCTGAGACAGAGCGGAGAAAGACAGGAGTTCGTTACT 4560
 QY 4561 GCCCCCACTGCTCGAGCGCTTCAATTTCTGAAAAGAGCTCGACTGGCGACCGGG 4620
 Db 4561 GCCCCCACTGCTCGAGCGCTTCAATTTCTGAAAAGAGCTCGACTGGCGACCGGG 4620
 QY 4621 CGGACTGTGCGCGGTGAGCATCACTTGGCAACCCCATCTGTGGCGACGAGCAAGGTT 4680


```

DB 4621 CCGAGTGTGCGGCTGCGATACCTTGGGCAACCCATCTCGCCGACGAGCAGTT 4680
QY 4681 GCGCGGAGGACCAAGCGCGGACCTTTCCTGCGCTTCCAGTCTGCGCTGCTGTC 4740
DB 4681 GCGCGGAGGACCAAGCGCGGACCTTTCCTGCGCTTCCAGTCTGCGCTGCTGTC 4740
QY 4741 TCGAAGCGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4800
DB 4741 TCGAAGCGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4800
QY 4801 TCGAAGCGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4860
DB 4801 TCGAAGCGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4860
QY 4861 GTCCAGGCTCTGATCAGCGGCGGCTTATCTCAGAGAGCGAGGCTTCTCCAGATCC 4920
DB 4861 GTCCAGGCTCTGATCAGCGGCGGCTTATCTCAGAGAGCGAGGCTTCTCCAGATCC 4920
QY 4921 AAGGCTACTCAAGAACTAAGGAGTGTCAATCCACTCTTTTCTGCGCGAGTGC 4980
DB 4921 AAGGCTACTCAAGAACTAAGGAGTGTCAATCCACTCTTTTCTGCGCGAGTGC 4980
QY 4981 GATTACAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5040
DB 4981 GATTACAGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5040
QY 5041 CGCTCATGATTCACAGAGCGGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 5100
DB 5041 CGCTCATGATTCACAGAGCGGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 5100
QY 5101 CGCTCATGATTCACAGAGCGGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 5160
DB 5101 CGCTCATGATTCACAGAGCGGCTTCTGCGCGGCGGCGGCGGCGGCGGCGGCGG 5160
QY 5161 ATTAGTTCTCTGTTGTTAAAGAGGCTGCTCCCATTTGAGTAAAGCTGCTGAG 5220
DB 5161 ATTAGTTCTCTGTTGTTAAAGAGGCTGCTCCCATTTGAGTAAAGCTGCTGAG 5220
QY 5221 TCCTCAAGTTCATCTGTTGAGTGTGCTGCGGTTGATTCATTCAGAGGAGG 5280
DB 5221 TCCTCAAGTTCATCTGTTGAGTGTGCTGCGGTTGATTCATTCAGAGGAGG 5280
QY 5281 GCGCAACTAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5312
DB 5281 GCGCAACTAGTACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 5312

```

```

RESULT 3
US-08-485-355B-43
; Sequence 43, Application US/08485355B
; Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hobbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 43:
SEQUENCE CHARACTERISTICS:
LENGTH: 5312 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 4518..4937
SEQUENCE DESCRIPTION: SEQ ID NO: 43:
US-08-485-355B-43
Query Match 100.0%; Score 5312; DB 3; Length 5312;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 5312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTTTCGCTCCCGCGGAGCGTAAATATAGGGGAAACAATGTACGGAAGGACAGAGT 60
DB 1 GTTTCGCTCCCGCGGAGCGTAAATATAGGGGAAACAATGTACGGAAGGACAGAGT 60
QY 61 GCGCGTGTCTAGCGCGCGGCAAGTGTGCTGAGCGGAAGTACTGACAGAGAGAGT 120
DB 61 GCGCGTGTCTAGCGCGCGGCAAGTGTGCTGAGCGGAAGTACTGACAGAGAGAGT 120
QY 121 AAGTTGACCTTGCAGCGCGGCACTGAGAGCACTGAAACCTCCAGACAGTGTACT 180
DB 121 AAGTTGACCTTGCAGCGCGGCACTGAGAGCACTGAAACCTCCAGACAGTGTACT 180
QY 181 GTGCGCTTCAAGAGGGGCACTTTACCCCGACCAACACCCGATCTGCGCGGCA 240
DB 181 GTGCGCTTCAAGAGGGGCACTTTACCCCGACCAACACCCGATCTGCGCGGCA 240
QY 241 CGTGTGCGAGAGAGGTTCTGCAATTTGCGCAAGGAGAGTGTGCAAGTGTGAG 300
DB 241 CGTGTGCGAGAGAGGTTCTGCAATTTGCGCAAGGAGAGTGTGCAAGTGTGAG 300
QY 301 GGGCGGCTCTGCAAGCGGCACTTAAGCTACATGAGGGGACCGAAGCGCGCGT 360
DB 301 GGGCGGCTCTGCAAGCGGCACTTAAGCTACATGAGGGGACCGAAGCGCGCGT 360
QY 361 TATCAGGGGTGACCAAGTACGAGCACCGGAGCGGCTCGGACACATTAAGGCT 420
DB 361 TATCAGGGGTGACCAAGTACGAGCACCGGAGCGGCTCGGACACATTAAGGCT 420
QY 421 TCTAGATCCGTGCGCACAGGCGCGGCGGAGTTCAAGCGGAGCGCTCACTG 480
DB 421 TCTAGATCCGTGCGCACAGGCGCGGCGGAGTTCAAGCGGAGCGCTCACTG 480
QY 481 GGCATTGCTCCCGGACCTTCTGCGTGAAGAGAGTGGGCTTGGCGGCTCAAT 540
DB 481 GGCATTGCTCCCGGACCTTCTGCGTGAAGAGAGTGGGCTTGGCGGCTCAAT 540
QY 541 GTTGAATGGCAATCACTCTCTATGAGAGTACCTTAGAGAGGCTGAGCAAT 600
DB 541 GTTGAATGGCAATCACTCTCTATGAGAGTACCTTAGAGAGGCTGAGCAAT 600
QY 601 GAGAACCGAGACTTCAATGTGTCGCGGCTTATGACATGCAATGCCAAGAGCT 660

```



```

Db      601 GAGAACCAAGGACTTTCATGATGTCGCGCTTTCATGACATGCAAGAGAGCTCTTAC 660
Qy      661 ATGACAAAGCTGTTAAATGCCAGCTCGGCTACCGCTTCCAGTTATTAAGAGCTATG 720
Db      661 ATGACAAAGCTGTTAAATGCCAGCTCGGCTACCGCTTCCAGTTATTAAGAGAGCTATG 720
Qy      721 GCTGTGAAGAGCTGCGATTCAGAGGGGGAGAGCTCCGCTCTCCACTTCCCTGAATTGAC 780
Db      721 GCTGTGAAGAGCTGCGATTCAGAGGGGGAGAGCTCCGCTCTCCACTTCCCTGAATTGAC 780
Qy      781 TTTCATCAACGAGAGCCAAAGAGCGGAGCATCGAGAGGCTGGCGCGCGGCTCTTACTCC 840
Db      781 TTTCATCAACGAGAGCCAAAGAGCGGAGCATCGAGAGGCTGGCGCGCGGCTCTTACTCC 840
Qy      841 AGACGCGCGCTCATTTTCTCCGCGAGAGCATGCGGGTGAATGCGTACTTACAGACTTC 900
Db      841 AGACGCGCGCTCATTTTCTCCGCGAGAGCATGCGGGTGAATGCGTACTTACAGACTTC 900
Qy      901 CACACATGAGCTCGGCTACTACTGTGTAGAGAACTAACCCACTCCGTTTGTCTTCACTC 960
Db      901 CACACATGAGCTCGGCTACTACTGTGTAGAGAACTAACCCACTCCGTTTGTCTTCACTC 960
Qy      961 CATATAGAAGTCCAGAGCGCGCAAGCTCCAGCATTTGAGCTGCGCATCACTCGCGCGCA 1020
Db      961 CATATAGAAGTCCAGAGCGCGCAAGCTCCAGCATTTGAGCTGCGCATCACTCGCGCGCA 1020
Qy      1021 CTTGAGAGCGGATGCTGCGCGCTGCTGCCAGAGAGCTGCCAAGGCTCTTGAGAAATCCA 1080
Db      1021 CTTGAGAGCGGATGCTGCGCGCTGCTGCCAGAGAGCTGCCAAGGCTCTTGAGAAATCCA 1080
Qy      1081 AACATCTTTTATTAAGCGCGAGAGCGCTCGGAGCATGAGCATTAAGCATCTTAACTCAG 1140
Db      1081 AACATCTTTTATTAAGCGCGAGAGCGCTCGGAGCATGAGCATTAAGCATCTTAACTCAG 1140
Qy      1141 CACAAAGTCAACATGCTGCTCAATTTTATGCAAAAGCGCTCTGAGAAAGAACTAGTCAG 1200
Db      1141 CACAAAGTCAACATGCTGCTCAATTTTATGCAAAAGCGCTCTGAGAAAGAACTAGTCAG 1200
Qy      1201 ATGACCGCTTGTATGCTGCTGCGCGCGCTAGAGCTGCGCGAGTGTGTGCTCAGAA 1260
Db      1201 ATGACCGCTTGTATGCTGCTGCGCGCGCTAGAGCTGCGCGAGTGTGTGCTCAGAA 1260
Qy      1261 GTCAACGAGAGCTCTGGAACATCTCACCGGCTGACCTGTGCGGACATGCTGTCTTT 1320
Db      1261 GTCAACGAGAGCTCTGGAACATCTCACCGGCTGACCTGTGCGGACATGCTGTCTTT 1320
Qy      1321 TACGTCCTCCACATCATGAGAGCGCGAGAGGCTGCGCTGCTCAAGACCGCCAGAGAC 1380
Db      1321 TACGTCCTCCACATCATGAGAGCGCGAGAGGCTGCGCTGCTCAAGACCGCCAGAGAC 1380
Qy      1381 GACGCTTTTGAAGAGACTTGTCTGGAAGAGTCTCAAGACAGTCTTGAGGCTCTGTTGC 1440
Db      1381 GACGCTTTTGAAGAGACTTGTCTGGAAGAGTCTCAAGACAGTCTTGAGGCTCTGTTGC 1440
Qy      1441 GGTCTGCGCAACCTTCAAGAGGACCGAGCGTCTTTTAAAGCGGTGTGATTAAGTAC 1500
Db      1441 GGTCTGCGCAACCTTCAAGAGGACCGAGCGTCTTTTAAAGCGGTGTGATTAAGTAC 1500
Qy      1501 CGAGTCCACTGCTCGAGAGATTAATCTGCAAGCTGCGCTGCTCCCTGGAACAGGTCCGC 1560
Db      1501 CGAGTCCACTGCTCGAGAGATTAATCTGCAAGCTGCGCTGCTCCCTGGAACAGGTCCGC 1560
Qy      1561 TTCTGCGCTCCCGGTAACACCTGCGCGCTCTTTCAGACAGAGGAAGAGCTTGAAGTTC 1620
Db      1561 TTCTGCGCTCCCGGTAACACCTGCGCGCTCTTTCAGACAGAGGAAGAGCTTGAAGTTC 1620
Qy      1621 CTTGCGGAAGCTGTGTGTAAGAGAGCTGCGGATCTTCCACTCCCTCTGTGAGAGAG 1680
Db      1621 CTTGCGGAAGCTGTGTGTAAGAGAGCTGCGGATCTTCCACTCCCTCTGTGAGAGAG 1680
Qy      1681 CCCCAAGGTTTCAAGCGCGAGCTTGTGAGACGCGAGCGGCGCTCACTCCCGAGTACCGC 1740
Db      1681 CCCCAAGGTTTCAAGCGCGAGCTTGTGAGACGCGAGCGGCGCTCACTCCCGAGTACCGC 1740
Qy      1741 GCCACTTTGAGAGAGGTCTTCAACACCGAGCTCAAGAGCTCAAGATCACCTCTGAGAAC 1800
Db      1741 GCCACTTTGAGAGAGGTCTTCAACACCGAGCTCAAGAGCTCAAGATCACCTCTGAGAAC 1800
Qy      1801 GCCCTCAAGACATTCAGAGGCTTACCTTCCCTGAGAGGCTCTGAGAGTATGAG 1860
Db      1801 GCCCTCAAGACATTCAGAGGCTTACCTTCCCTGAGAGGCTCTGAGAGTATGAG 1860
Qy      1861 GGCCTCGCAGAGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Db      1861 GGCCTCGCAGAGCAGCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1920
Qy      1921 AAGGACTTTTACGTGAGACCCACCAAGAGAACTGAGAGAGGCTATGAGCCGCGATCAA 1980
Db      1921 AAGGACTTTTACGTGAGACCCACCAAGAGAACTGAGAGAGGCTATGAGCCGCGATCAA 1980
Qy      1981 CCGCGCTCGGCTCGGCTACAGCAATGTGCGCTTGCAGATTCTCCGCTGTGCAACGCC 2040
Db      1981 CCGCGCTCGGCTCGGCTACAGCAATGTGCGCTTGCAGATTCTCCGCTGTGCAACGCC 2040
Qy      2041 GAGGCGCGCCCTTTCGCTTACCGTGTATGAGAGAGTCTTCAATGTTCCGCTGTGTAC 2100
Db      2041 GAGGCGCGCCCTTTCGCTTACCGTGTATGAGAGAGTCTTCAATGTTCCGCTGTGTAC 2100
Qy      2101 GTGCGATCGTGACCGCTGTGTGTGCGGAGCTCAAGAAATGCTTGTGAGAGAGTCCAC 2160
Db      2101 GTGCGATCGTGACCGCTGTGTGTGCGGAGCTCAAGAAATGCTTGTGAGAGAGTCCAC 2160
Qy      2161 CAAATCGGTTTATGAGCTTCAAGAGCAAGAGCGAGCAATGCGGCTGTTGCGAGGTC 2220
Db      2161 CAAATCGGTTTATGAGCTTCAAGAGCAAGAGCGAGCAATGCGGCTGTTGCGAGGTC 2220
Qy      2221 GTTAAAGAGTGCCTGCGGCGCACTTTCAACCAACCAAGCGTGTGCGGCGAGGTCGTT 2280
Db      2221 GTTAAAGAGTGCCTGCGGCGCACTTTCAACCAACCAAGCGTGTGCGGCGAGGTCGTT 2280
Qy      2281 GCCACACGTTTTTCCAGAGCTTGTATCCCGGGTGTGCAACCACTCAAGGTGTGTGCA 2340
Db      2281 GCCACACGTTTTTCCAGAGCTTGTATCCCGGGTGTGCAACCACTCAAGGTGTGTGCA 2340
Qy      2341 TCCATGAGCAGCGTGCAGGAGCTACCGGCAAGAGGCGCAAGAGCTCTGCTTCAAG 2400
Db      2341 TCCATGAGCAGCGTGCAGGAGCTACCGGCAAGAGGCGCAAGAGCTCTGCTTCAAG 2400
Qy      2401 CAGGAGAGAAAGTGCAGCAAGGAGCTGAGAGCGGATGACTGTGACGAAGCGCAGAGGA 2460
Db      2401 CAGGAGAGAAAGTGCAGCAAGGAGCTGAGAGCGGATGACTGTGACGAAGCGCAGAGGA 2460
Qy      2461 CGCACTTTTGTGTGTATTTCTGATTTTCAAGAGGCTTCAAGAGAGAGAGAGAGTCTTC 2520
Db      2461 CGCACTTTTGTGTGTATTTCTGATTTTCAAGAGGCTTCAAGAGAGAGAGAGAGTCTTC 2520
Qy      2521 GCTGAGAGTGCAGCTTCTAGTGCAGTCAAGGCGCAACCAACCACTGTATCATCCGC 2580
Db      2521 GCTGAGAGTGCAGCTTCTAGTGCAGTCAAGGCGCAACCAACCACTGTATCATCCGC 2580
Qy      2581 GACCCGACAGGTGACATTTGAGAGCAACTCAACATAGCGGAAAGCGAGGTGTTTCA 2640
Db      2581 GACCCGACAGGTGACATTTGAGAGCAACTCAACATAGCGGAAAGCGAGGTGTTTCA 2640
Qy      2641 GACATCCCTGTGACCCCTGAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAG 2700
Db      2641 GACATCCCTGTGACCCCTGAGAGTCAAGAGTCAAGAGTCAAGAGTCAAGAGTCAAG 2700
Qy      2701 GAAAGTATGAGCAAGTACCCCGCAGAGTGCAGCGCGCAGAGCAATCATCTGCTC 2760
Db      2701 GAAAGTATGAGCAAGTACCCCGCAGAGTGCAGCGCGCAGAGCAATCATCTGCTC 2760
Qy      2761 CGCAAGAACTTTCGAGAGCAACCCGAGCTGTGTGTGCTTGTGCGAAGACGCGTAC 2820
Db      2761 CGCAAGAACTTTCGAGAGCAACCCGAGCTGTGTGTGCTTGTGCGAAGACGCGTAC 2820

```

2821 GAGGTGTTGGCGGTGTCGCAAAATCAACGTAGAGCTTGCCGAACCCGACGCGACCCG 2880
2821 GAGGTGTTGGCGGTGTCGTCGCAAAATCAACGTAGAGCTTGCCGAACCCGACGCGACCCG 2880
2881 AAGCCGATAGGGCGTTCAGAGAAAGGGGTACAGTGGGTCAAGGTCAACCAACGCGTCTAAC 2940
2881 AAGCCGATAGGGCGTTCAGAGAAAGGGGTACAGTGGGTCAAGGTCAACCAACGCGTCTAAC 2940
2941 AAACACGAGCGGTCCAGACGCTGTGTGTCGCTGACCAACAGGAAAGGCTGACCTGCGG 3000
2941 AAACACGAGCGGTCCAGACGCTGTGTGTCGCTGACCAACAGGAAAGGCTGACCTGCGG 3000
3001 CTACACGAAGCTAAGAGAGACGTCAAAACGATGCTAAACTCGCTTGAACGCAATTGGAGAC 3060
3001 CTACACGAAGCTAAGAGAGACGTCAAAACGATGCTAAACTCGCTTGAACGCAATTGGAGAC 3060
3061 TGGACTGTCACTGAAGACGCGCGGTGACCGAGCTGTCTTGAGAACCCAGCTCAAGTTCAAC 3120
3061 TGGACTGTCACTGAAGACGCGCGGTGACCGAGCTGTCTTGAGAACCCAGCTCAAGTTCAAC 3120
3121 CAACGCGCGGACCGTGAAGAACTGCTGAGACGACGACCCCTTACATCCGTGACATA 3180
3121 CAACGCGCGGACCGTGAAGAACTGCTGAGACGACGACCCCTTACATCCGTGACATA 3180
3181 GACCTTCCTTATGAGACTGAGAGAAAGTGTGCGCCAGGCGGATCAATACGCGGAAAGTTC 3240
3181 GACCTTCCTTATGAGACTGAGAGAAAGTGTGCGCCAGGCGGATCAATACGCGGAAAGTTC 3240
3241 GGGCAGGGGATGCGCGCTCACTCAAAAGTCTCTCAACTTCGCTGCGCGCTTGGATACGC 3300
3241 GGGCAGGGGATGCGCGCTCACTCAAAAGTCTCTCAACTTCGCTGCGCGCTTGGATACGC 3300
3301 ATACTCGAGAGATCTCCGTACCGGAGCCGACGCGTCCGGTACAGCAACGCTCTCCGC 3360
3301 ATACTCGAGAGATCTCCGTACCGGAGCCGACGCGTCCGGTACAGCAACGCTCTCCGC 3360
3361 GACGAAAGAGGGCCATGCTGTCGAAAGGAGATCAATCAAGTCCCAACGCGCAAGTTTC 3420
3361 GACGAAAGAGGGCCATGCTGTCGAAAGGAGATCAATCAAGTCCCAACGCGCAAGTTTC 3420
3421 GTCTCGGCGGACTGACCGAGTTTGAACCGCCCAAAATTAACAGAGTGAAGTGTCTTTC 3480
3421 GTCTCGGCGGACTGACCGAGTTTGAACCGCCCAAAATTAACAGAGTGAAGTGTCTTTC 3480
3481 GCCGCCCTTTTGAAGCGCATCGGCAAGCTGAGCTGCCGTTAATCTATTTCAGAGAACGG 3540
3481 GCCGCCCTTTTGAAGCGCATCGGCAAGCTGAGCTGCCGTTAATCTATTTCAGAGAACGG 3540
3541 TGTGGGAAACGCACTTGGAGACGAAAGGTCTAAGCTCCGTTGAAGTGAAGTGTCTGCTC 3600
3541 TGTGGGAAACGCACTTGGAGACGAAAGGTCTAAGCTCCGTTGAAGTGAAGTGTCTGCTC 3600
3601 GACTCGGCGGAGCTTGAACGCTTGCAGCAACCATCTTCTGCGCGCGCTCATGCTC 3660
3601 GACTCGGCGGAGCTTGAACGCTTGCAGCAACCATCTTCTGCGCGCGCTCATGCTC 3660
3661 ACGCTCTTCGCGGAGCTGAAGTTGCAAGCTTTCAAAAGCGAGCATCGCTCTCTGTGGT 3720
3661 ACGCTCTTCGCGGAGCTGAAGTTGCAAGCTTTCAAAAGCGAGCATCGCTCTCTGTGGT 3720
3721 AGCCATTACCTCCGTTTGAACGCTAGCGCTTTCACATGGGGGAAAGTTAAAGAACAA 3780
3721 AGCCATTACCTCCGTTTGAACGCTAGCGCTTTCACATGGGGGAAAGTTAAAGAACAA 3780
3781 CATTGAAGGTGAGAGTGAAGAAATCGTCCGTACATGAGACTCTCGTCTCCGCTAG 3840
3781 CATTGAAGGTGAGAGTGAAGAAATCGTCCGTACATGAGACTCTCGTCTCCGCTAG 3840
3841 CAGGTGCTCTGACCTGTCAAGAGCGCTTCAAGATATTTGGCGCTGCTAACAAAGC 3900
3841 CAGGTGCTCTGACCTGTCAAGAGCGCTTCAAGATATTTGGCGCTGCTAACAAAGC 3900
3901 GAACCTCTTTACTCCAAAGTACGTGAGGCTGTGAGAGACATCAACCAAGGCTGAGTAC 3960
3901 GAACCTCTTTACTCCAAAGTACGTGAGGCTGTGAGAGACATCAACCAAGGCTGAGTAC 3960
3961 GCCGCTACCAAGCTCTCTGTGCAATGTACATCTTACATTAATTAAGCGCGGAG 4020
3961 GCCGCTACCAAGCTCTCTGTGCAATGTACATCTTACATTAATTAAGCGCGGAG 4020
4021 TCTGGGCGGTACATCATGAGCGCTGTGTGCTTGGGCGCGGAGCTTCCGTTTAA 4080
4021 TCTGGGCGGTACATCATGAGCGCTGTGTGCTTGGGCGCGGAGCTTCCGTTTAA 4080
4081 CAACCTGCGGTGTCGTCGTCATGTGAGGCAACCGGCTTACAGAGACATGATCCG 4140
4081 CAACCTGCGGTGTCGTCGTCATGTGAGGCAACCGGCTTACAGAGACATGATCCG 4140
4081 CAACCTGCGGTGTCGTCGTCATGTGAGGCAACCGGCTTACAGAGACATGATCCG 4140
4081 CAACCTGCGGTGTCGTCGTCATGTGAGGCAACCGGCTTACAGAGACATGATCCG 4140
4141 GCTAACGTGCGGATGCTGCTTGAACAGCTTTCGAGCCCGGACGCGCGCGCGCG 4200
4141 GCTAACGTGCGGATGCTGCTTGAACAGCTTTCGAGCCCGGACGCGCGCGCGCGCGCG 4200
4141 GCTAACGTGCGGATGCTGCTTGAACAGCTTTCGAGCCCGGACGCGCGCGCGCGCGCG 4200
4141 GCTAACGTGCGGATGCTGCTTGAACAGCTTTCGAGCCCGGACGCGCGCGCGCGCGCG 4200
4201 GCAAGTTTCGTTGCGGACATGTGCGAAAGCGGAAAGCCCTTCTTACCTTACCGGAAAGCT 4260
4201 GCAAGTTTCGTTGCGGACATGTGCGAAAGCGGAAAGCCCTTCTTACCTTACCGGAAAGCT 4260
4201 GCAAGTTTCGTTGCGGACATGTGCGAAAGCGGAAAGCCCTTCTTACCTTACCGGAAAGCT 4260
4201 GCAAGTTTCGTTGCGGACATGTGCGAAAGCGGAAAGCCCTTCTTACCTTACCGGAAAGCT 4260
4261 GGTGTTTCTGCACTACAGGCAAGTTCGAGCTGAGGACCTGCGCGCGGAGTCTCATG 4320
4261 GGTGTTTCTGCACTACAGGCAAGTTCGAGCTGAGGACCTGCGCGCGGAGTCTCATG 4320
4261 GGTGTTTCTGCACTACAGGCAAGTTCGAGCTGAGGACCTGCGCGCGGAGTCTCATG 4320
4261 GGTGTTTCTGCACTACAGGCAAGTTCGAGCTGAGGACCTGCGCGCGGAGTCTCATG 4320
4321 GATGCACTGCAAGCAAGCTTTTCGAGTTTATTAACACCGGAGACCCCGTCCATCA 4380
4321 GATGCACTGCAAGCAAGCTTTTCGAGTTTATTAACACCGGAGACCCCGTCCATCA 4380
4321 GATGCACTGCAAGCAAGCTTTTCGAGTTTATTAACACCGGAGACCCCGTCCATCA 4380
4321 GATGCACTGCAAGCAAGCTTTTCGAGTTTATTAACACCGGAGACCCCGTCCATCA 4380
4381 TCTGCGCGGTATGCTTTCATTCGAGCTCTCTTACATGTGTGAAAGTCTCATGAGT 4440
4381 TCTGCGCGGTATGCTTTCATTCGAGCTCTCTTACATGTGTGAAAGTCTCATGAGT 4440
4381 TCTGCGCGGTATGCTTTCATTCGAGCTCTCTTACATGTGTGAAAGTCTCATGAGT 4440
4381 TCTGCGCGGTATGCTTTCATTCGAGCTCTCTTACATGTGTGAAAGTCTCATGAGT 4440
4441 GAGAGACCGGAAAGGACCAAGAAAGCTTGAACAGCAGCAAGGCGCTTCCGAAAGAG 4500
4441 GAGAGACCGGAAAGGACCAAGAAAGCTTGAACAGCAGCAAGGCGCTTCCGAAAGAG 4500
4441 GAGAGACCGGAAAGGACCAAGAAAGCTTGAACAGCAGCAAGGCGCTTCCGAAAGAG 4500
4441 GAGAGACCGGAAAGGACCAAGAAAGCTTGAACAGCAGCAAGGCGCTTCCGAAAGAG 4500
4501 CAATCAGCTCGTCTGAATGTCTGACAGAAAGCGGAAAGGACGAGTTCGTTAACT 4560
4501 CAATCAGCTCGTCTGAATGTCTGACAGAAAGCGGAAAGGACGAGTTCGTTAACT 4560
4501 CAATCAGCTCGTCTGAATGTCTGACAGAAAGCGGAAAGGACGAGTTCGTTAACT 4560
4501 CAATCAGCTCGTCTGAATGTCTGACAGAAAGCGGAAAGGACGAGTTCGTTAACT 4560
4561 GCCCCTCATGCTCGGAGCGCTCATCTTTCATTTTCGAAAGAGCTGACGCGGACGCGG 4620
4561 GCCCCTCATGCTCGGAGCGCTCATCTTTCATTTTCGAAAGAGCTGACGCGGACGCGG 4620
4561 GCCCCTCATGCTCGGAGCGCTCATCTTTCATTTTCGAAAGAGCTGACGCGGACGCGG 4620
4561 GCCCCTCATGCTCGGAGCGCTCATCTTTCATTTTCGAAAGAGCTGACGCGGACGCGG 4620
4621 CGAAGTGTGCGGCTGCGGACATCACTTCGAGCAACCCATCCTGCGGACGAGGTT 4680
4621 CGAAGTGTGCGGCTGCGGACATCACTTCGAGCAACCCATCCTGCGGACGAGGTT 4680
4621 CGAAGTGTGCGGCTGCGGACATCACTTCGAGCAACCCATCCTGCGGACGAGGTT 4680
4621 CGAAGTGTGCGGCTGCGGACATCACTTCGAGCAACCCATCCTGCGGACGAGGTT 4680
4681 GCCGAGAGACCAAGCGGACTTTGCGCTTTCCTGAGTCCAGTCTGCGCGCTGTC 4740
4681 GCCGAGAGACCAAGCGGACTTTGCGCTTTCCTGAGTCCAGTCTGCGCGCTGTC 4740
4681 GCCGAGAGACCAAGCGGACTTTGCGCTTTCCTGAGTCCAGTCTGCGCGCTGTC 4740
4681 GCCGAGAGACCAAGCGGACTTTGCGCTTTCCTGAGTCCAGTCTGCGCGCTGTC 4740
4681 GCCGAGAGACCAAGCGGACTTTGCGCTTTCCTGAGTCCAGTCTGCGCGCTGTC 4740
4741 TCGAAGCGGTACCGGCGCGGCGGACGCTGCGGTGAAAGAGTCAACCCCGCTCCACGCG 4800
4741 TCGAAGCGGTACCGGCGCGGCGGACGCTGCGGTGAAAGAGTCAACCCCGCTCCACGCG 4800
4741 TCGAAGCGGTACCGGCGCGGCGGACGCTGCGGTGAAAGAGTCAACCCCGCTCCACGCG 4800
4741 TCGAAGCGGTACCGGCGCGGCGGACGCTGCGGTGAAAGAGTCAACCCCGCTCCACGCG 4800
4801 TGGAAGGGCGTGAACCGGAGACGACCGGAGTCAAGGGAGGACCGGAGACAGCGCGGCTC 4860
4801 TGGAAGGGCGTGAACCGGAGACGACCGGAGTCAAGGGAGGACCGGAGACAGCGCGGCTC 4860
4801 TGGAAGGGCGTGAACCGGAGACGACCGGAGTCAAGGGAGGACCGGAGACAGCGCGGCTC 4860
4801 TGGAAGGGCGTGAACCGGAGACGACCGGAGTCAAGGGAGGACCGGAGACAGCGCGGCTC 4860
4801 TGGAAGGGCGTGAACCGGAGACGACCGGAGTCAAGGGAGGACCGGAGACAGCGCGGCTC 4860
4861 GTCCAGGCTGTGATCAGCGCGCGTATCTGAGAGAGAAAGCTTCTCCGACGCAATCC 4920
4861 GTCCAGGCTGTGATCAGCGCGCGTATCTGAGAGAGAAAGCTTCTCCGACGCAATCC 4920
4861 GTCCAGGCTGTGATCAGCGCGCGTATCTGAGAGAGAAAGCTTCTCCGACGCAATCC 4920
4861 GTCCAGGCTGTGATCAGCGCGCGTATCTGAGAGAGAAAGCTTCTCCGACGCAATCC 4920
4921 AAAGGCTACTCAAGAACTTAAGAGATGCTCAATCACTCTTTCCTGCGCGGAGTGG 4980
4921 AAAGGCTACTCAAGAACTTAAGAGATGCTCAATCACTCTTTCCTGCGCGGAGTGG 4980
4921 AAAGGCTACTCAAGAACTTAAGAGATGCTCAATCACTCTTTCCTGCGCGGAGTGG 4980
4921 AAAGGCTACTCAAGAACTTAAGAGATGCTCAATCACTCTTTCCTGCGCGGAGTGG 4980
4981 GATTACAGGCGCGGACCTGCGAAGCAAGTCTGCGCGCGGCTGAGAGATGGG 5040
4981 GATTACAGGCGCGGACCTGCGAAGCAAGTCTGCGCGCGGCTGAGAGATGGG 5040

```

Db      4981 GATTACAGGCGCGGCGAGCTCCGAGAGTGTGCCGCGCGCTGCAGAGATGCGG 5040
Qy      5041 CGCTCATGATTTACAGAGCGCGTGGCTTCAATCTGCGCGCGAGTCCGACTTGAAGCGGATA 5100
Db      5041 CGCTCATGATTTACAGAGCGCGTGGCTTCAATCTGCGCGCGAGTCCGACTTGAAGCGGATA 5100
Qy      5101 CGCTCAACCTCGGAGCTGTGTCCGATGTAAAGATCAGAAAGCGCATGAAGAAACAA 5160
Db      5101 CGCTCAACCTCGGAGCTGTGTCCGATGTAAAGATCAGAAAGCGCATGAAGAAACAA 5160
Qy      5161 ATTAGTTTCTTGTTCGTAACAAGAGTGTCCCTCCCATTTAGGTAAAGACTGTGTAG 5220
Db      5161 ATTAGTTTCTTGTTCGTAACAAGAGTGTCCCTCCCATTTAGGTAAAGACTGTGTAG 5220
Qy      5221 TCCTCAACGTTACTGTTAGTGTGTGCGCGTTCGATTCATTCGCAAGCAAGAGGT 5280
Db      5221 TCCTCAACGTTACTGTTAGTGTGTGCGCGTTCGATTCATTCGCAAGCAAGAGGT 5280
Qy      5281 GCGCACTAGTACGGCGCGCGCTGGGATACCA 5312
Db      5281 GCGCACTAGTACGGCGCGCGCTGGGATACCA 5312

```

RESULT 4
US-08-485-355B-45
Sequence 45: Application US/08485355B
Patent No. 6177075

GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H. J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants

NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hobach Test Albitton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992

ATTORNEY/AGENT INFORMATION:
NAME: Treceartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 5368 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS

LOCATION: 4944..5162
SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-08-485-355B-45

Query Match 98.0%, Score 5206, DB 3, Length 5368,
Best Local Similarly 99.0%, Pred. No. 0,
Matches 5312, Conservative 0, Mismatches 0, Indels 56, Gaps 5;

```

Qy      1 GTTCTGCTCCCGGAGCGGTAAATATAGGGGAACAATGACGGAAGGACAGAGCGTG 60
Db      1 GTTCTGCTCCCGGAGCGGTAAATATAGGGGAACAATGACGGAAGGACAGAGCGTG 60
Qy      61 GCGCGTGTACGCGCGGAGATGTGCGCTACGCGAAAGTACTGACAGAGAGAGTGC 120
Db      61 GCGCGTGTACGCGCGGAGATGTGCGCTACGCGAAAGTACTGACAGAGAGAGTGC 120
Qy      121 AAGTTGAGCTTGCGCGCGGACCTGAAGGACCTAGAAACCTCCACAGACTGTACTATCCG 180
Db      121 AAGTTGAGCTTGCGCGCGGACCTGAAGGACCTAGAAACCTCCACAGACTGTACTATCCG 180
Qy      181 CTGCGCTTCAAGGGGGGCACTTAAACCCCGAGCAACAACCCGATCTGTGGCGGGGACCA 240
Db      181 CTGCGCTTCAAGGGGGGCACTTAAACCCCGAGCAACAACCCGATCTGTGGCGGGGACCA 240
Qy      241 CGTGTGCGAGAAAGGTTCTGCACAATTTGCGCGAGGGAGCTAGACAGTGTCTGAGATA 300
Db      241 CGTGTGCGAGAAAGGTTCTGCACAATTTGCGCGAGGGAGCTAGACAGTGTCTGAGATA 300
Qy      301 GGGCGGTCTTGTGACAGCGGACCTTAAGCTACATGGGGACCGAAACGCGCGGTGCGAGAC 360
Db      301 GGGCGGTCTTGTGACAGCGGACCTTAAGCTACATGGGGACCGAAACGCGCGGTGCGAGAC 360
Qy      361 TATCAGGGGTGACCAAGTACGCGGACCGCGGAGCGCTGCGGACACATTAAGGCTTGTAG 420
Db      361 TATCAGGGGTGACCAAGTACGCGGACCGCGGAGCGCTGCGGACACATTAAGGCTTGTAG 420
Qy      421 TATCAGGGGTGACCAAGTACGCGGACCGCGGAGCGCTGCGGACACATTAAGGCTTGTAG 480
Db      421 TATCAGGGGTGACCAAGTACGCGGACCGCGGAGCGCTGCGGACACATTAAGGCTTGTAG 480
Qy      481 GGCATTGCTCCCGGACCTTGTGCGTGTGACAGAGTGTGCGCTTGTGCGGCTTGTGCGG 540
Db      481 GGCATTGCTCCCGGACCTTGTGCGTGTGACAGAGTGTGCGCTTGTGCGGCTTGTGCGG 540
Qy      541 GTTGAATTGCGAATACCTCCCTTATGAGTACCTTAGAGAGGCTGGCAATGCGTTT 600
Db      541 GTTGAATTGCGAATACCTCCCTTATGAGTACCTTAGAGAGGCTGGCAATGCGTTT 600
Qy      601 GAGAACCAAGGACCTTCAATGATGTCCGCGCTTATGACATGACATGCGAGAGAGTGTCTAC 660
Db      601 GAGAACCAAGGACCTTCAATGATGTCCGCGCTTATGACATGACATGCGAGAGAGTGTCTAC 660
Qy      661 ATGGAACAAGTGTAAATGCCAGAGCTGGCTACCGCTTCAAGTAAATGAGAGCTTATG 720
Db      661 ATGGAACAAGTGTAAATGCCAGAGCTGGCTACCGCTTCAAGTAAATGAGAGCTTATG 720
Qy      721 GCTGTGAAGGACTGCGGATTCGAGGGGGGAGACCTCCGCTCCACCTCCGAGTGTGAG 780
Db      721 GCTGTGAAGGACTGCGGATTCGAGGGGGGAGACCTCCGCTCCACCTCCGAGTGTGAG 780
Qy      781 TTCAATCAAGAGAGCCCAAGAGCGGCGATCGAGAGGCTGCGCGCGCTCTCTACTCC 840
Db      781 TTCAATCAAGAGAGCCCAAGAGCGGCGATCGAGAGGCTGCGCGCGCTCTCTACTCC 840
Qy      841 AGACGCGCGCTCATTTTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Db      841 AGACGCGCGCTCATTTTCTCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 900
Qy      901 CACACATGGCTCGCTTACTTACTGTAGAGAACTAACCCCACTCGTTTGTGTTTCTCACTC 960
Db      901 CACACATGGCTCGCTTACTTACTGTAGAGAACTAACCCCACTCGTTTGTGTTTCTCACTC 960
Qy      961 CATATGAAAGTCAAGAGCGGCGCAAGGCTTCAAGATGAGAGTGTGCGCATCACTGCGCGCA 1020

```


QY 3181 GACTTCCTTATGAGACTGACGAGAAAGTGTGCGCCCAAGCCGATCAATACGGGCAAGTTC 3240
DB 3181 GACTTCCTTATGAGACTGACGAGAAAGTGTGCGCCCAAGCCGATCAATACGGGCAAGTTC 3240
QY 3241 GGGCAGGAGATCGCGCTGACTCAAAAGTCTCTCAACTTCGCTCGCGCTTGGATACGC 3300
DB 3241 GGGCAGGAGATCGCGCTGACTCAAAAGTCTCTCAACTTCGCTCGCGCTTGGATACGC 3300
QY 3301 ATACTGAGAGATCTCCGTAACCGGAGCCGCACGGTCCGATCAGCAACGGTCTCCC 3360
DB 3301 ATACTGAGAGATCTCCGTAACCGGAGCCGCACGGTCCGATCAGCAACGGTCTCCC 3360
QY 3361 GACGAGAAAGAGCGCATGCTGCTCGAAGCAAGATCAATGATCCACACGCGACGCTTC 3420
DB 3361 GACGAGAAAGAGCGCATGCTGCTCGAAGCAAGATCAATGATCCACACGCGACGCTTC 3420
QY 3421 GTCTCGGCGGACTGAGACCGAGTTTGACACGCGCCCAATACAGATGAGCTGCTTC 3480
DB 3421 GTCTCGGCGGACTGAGACCGAGTTTGACACGCGCCCAATACAGATGAGCTGCTTC 3480
QY 3481 GCGGCGCTTTTAAAGCCGATCGACGCGCTGACGCTGCTTAATCTAATCAGAGAA 3540
DB 3481 GCGGCGCTTTTAAAGCCGATCGACGCGCTGACGCTGCTTAATCTAATCAGAGAA 3540
QY 3541 TGTTGGAAACGACCTTTGCGAGCGAAGGCTTACGCTCGTTGAAGTGCAGCTGCTC 3600
DB 3541 TGTTGGAAACGACCTTTGCGAGCGAAGGCTTACGCTCGTTGAAGTGCAGCTGCTC 3600
QY 3601 GACTCCGCGGAGTGTGAGAGCGCTTGCGGCAACCACTTCTGCGCGCGCTGATGCTC 3660
DB 3601 GACTCCGCGGAGTGTGAGAGCGCTTGCGGCAACCACTTCTGCGCGCGCTGATGCTC 3660
QY 3661 ACGCTCTTCGCGGCGCTCAAGTTGCGAGCTTTCAAAAGCGACGACTGCTCTGTGCT 3720
DB 3661 ACGCTCTTCGCGGCGCTCAAGTTGCGAGCTTTCAAAAGCGACGACTGCTCTGTGCT 3720
QY 3721 AGCCATTACCTCGTTTGCGAGCTGACGCGCTTCAATGCGGAAAGCTTCAAGACCAA 3780
DB 3721 AGCCATTACCTCGTTTGCGAGCTGACGCGCTTCAATGCGGAAAGCTTCAAGACCAA 3780
QY 3781 CATTTGAAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3840
DB 3781 CATTTGAAGTGTGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3840
QY 3841 CAGGTGCTCTGAGACCTGTCAGAGAGCGCTCTCAAGATATTTGGGCGCTGCTACAGAGC 3900
DB 3841 CAGGTGCTCTGAGACCTGTCAGAGAGCGCTCTCAAGATATTTGGGCGCTGCTACAGAGC 3900
QY 3901 GAATCTCTTATCTCAAGTACGTGAGAGGCTGTGAGAGGATACCAAGAGGCTGAGAGTAC 3960
DB 3901 GAATCTCTTATCTCAAGTACGTGAGAGGCTGTGAGAGGATACCAAGAGGCTGAGAGTAC 3960
QY 3961 GCGGCGTACCAACAGCTCTGTCGACATGTGACAGATGCTACTCAATTAAGCGCGGAG 4020
DB 3961 GCGGCGTACCAACAGCTCTGTCGACATGTGACAGATGCTACTCAATTAAGCGCGGAG 4020
QY 4021 TCTGCGGCTACATCATGACGCTGTGCTTGGGCGGCGGAGCTTCCGTTGAA 4080
DB 4021 TCTGCGGCTACATCATGACGCTGTGCTTGGGCGGCGGAGCTTCCGTTGAA 4080
QY 4081 CAATCGCGCTGTGAGCGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4140
DB 4081 CAATCGCGCTGTGAGCGCTGACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4140
QY 4141 GCTAACGTCGCGCTGTGAGCGCTTGAACAGCTTTCGAGCGCGCGCAGCGCGCGCG 4200
DB 4141 GCTAACGTCGCGCTGTGAGCGCTTGAACAGCTTTCGAGCGCGCGCAGCGCGCGCGCG 4200
QY 4201 GCAAGTTCTGTTGAGAGCTGTCGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4260
DB 4201 GCAAGTTCTGTTGAGAGCTGTCGAGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4260

QY 4261 GGTGTTTCTGCGACTCAAGCCGACTGTCGAGCTGCGGCGCGGAGTCTCATG 4320
DB 4261 GGTGTTTCTGCGACTCAAGCCGACTGTCGAGCTGCGGCGCGGAGTCTCATG 4320
QY 4321 GATGACCTGAGACCAAGCTTTTGGAGTTATGACACCGGAGACCCGTCACATCA 4380
DB 4321 GATGACCTGAGACCAAGCTTTTGGAGTTATGACACACCGGAGACCCGTCACATCA 4380
QY 4381 TCTTCGCGCTGATGCTGCTTCAATGAGCTCTCTCAATGCTGTGAGAGGAGTGTCA 4440
DB 4381 TCTTCGCGCTGATGCTGCTTCAATGAGCTCTCTCAATGCTGTGAGAGGAGTGTCA 4440
QY 4441 GAGACACCGCAAGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4500
DB 4441 GAGACACCGCAAGACCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4500
QY 4501 CAATCAAGCTGTGATGATGCTGAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4560
DB 4501 CAATCAAGCTGTGATGATGCTGAGACAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 4560
QY 4561 GCGGCGCTGCTCGAGCGCTTCAATCTCAATTTTGGAGGAGGAGGAGGAGGAGGAGGAG 4620
DB 4561 GCGGCGCTGCTCGAGCGCTTCAATCTCAATTTTGGAGGAGGAGGAGGAGGAGGAGGAG 4620
QY 4621 CCGACTGTGCGCGCTGCGACATGACCTTGGGCAACCCCATCTGCGCGACAGGAGT 4680
DB 4621 CCGACTGTGCGCGCTGCGACATGACCTTGGGCAACCCCATCTGCGCGACAGGAGT 4680
QY 4681 GCGCGAGAGCAAGCGCGGACTTTCGCGCTTTCCTGAGGTTCCAGTTCGCGCTGCTG 4740
DB 4681 GCGCGAGAGCAAGCGCGGACTTTCGCGCTTTCCTGAGGTTCCAGTTCGCGCTGCTG 4740
QY 4741 TCGAAGCGGATACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4800
DB 4741 TCGAAGCGGATACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4800
QY 4801 TCGAAGCGGATACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4860
DB 4801 TCGAAGCGGATACGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 4860
QY 4861 GTCGAGGCTGTATCAGCGCGCGCTTATCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 4920
DB 4861 GTCGAGGCTGTATCAGCGCGCGCTTATCTCAAGAGGAGGAGGAGGAGGAGGAGGAGGAG 4920
QY 4921 AAGGCTACTCAAGACTTAAGGAGTGTCAATCAATCCACTTTTCTGCTG---CCCGAGT 4977
DB 4921 AAGGCTACTCAAGACTTAAGGAGTGTCAATCAATCCACTTTTCTGCTG---CCCGAGT 4977
QY 4978 GCGGATTAACAGAGCG 5030
DB 4978 GCGGATTAACAGAGCG 5030
QY 5031 AGAGATGCGCGCTCATGATATTCAGAGCGCTTGGC-----TTTCAAT 5071
DB 5031 AGAGATGCGCGCTCATGATATTCAGAGCGCTTGGC-----TTTCAAT 5071
QY 5071 GTCGCGCGAGTGCAGCTTGAAGGCTCATGCTCTTACTCGGA-----CTCT 5118
DB 5071 GTCGCGCGAGTGCAGCTTGAAGGCTCATGCTCTTACTCGGA-----CTCT 5118
QY 5119 GTTCCCGATGTAAGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5160
DB 5119 GTTCCCGATGTAAGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5160
QY 5161 GTTCCCGATGTAAGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5220
DB 5161 GTTCCCGATGTAAGATCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5220
QY 5221 GTTTCCTGTTGCTGAACAAGTGTGCTCCATTTGAGGATTAAGAGCTGTGAGTGTCT 5280
DB 5221 GTTTCCTGTTGCTGAACAAGTGTGCTCCATTTGAGGATTAAGAGCTGTGAGTGTCT 5280
QY 5285 CAAGGTTACTGTTGAGAGTGTGCTGCGGTTCAATTCATTCAGAGCAAGCAAGGAGTGGC 5340
DB 5285 CAAGGTTACTGTTGAGAGTGTGCTGCGGTTCAATTCATTCAGAGCAAGCAAGGAGTGGC 5340

DB 5341 AACTAGTACGGCGCCCTGGGATACCA 5368

RESULT 5

US-09-194-613-4

Sequence 4, Application US/09194613

Patent No. 6251654

GENERAL INFORMATION:

APPLICANT: GORDON, Karl H.

APPLICANT: HANZLIK, Terry N.

TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES

NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:

ADDRESSEE: Medemott, Wall & Emery

STREET: 600 13th Street, N.W.

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20005-3096

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/194,613

FILING DATE: 30-NOV-1998

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Bucca Ph.D., Daniel

REGISTRATION NUMBER: 42,368

REFERENCE/DOCKET NUMBER: 50179-061

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-756-8000

TELEFAX: 202-756-8087

INFORMATION FOR SEQ ID NO: 4:

SEQUENCE CHARACTERISTICS:

LENGTH: 6534 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-09-194-613-4

Query Match 1.4%; Score 76.4; DB 3; Length 6534;
Best Local Similarity 52.5%; Pred. No. 5, 1e-09;
Matches 192; Conservative 0; Mismatches 171; Indels 3; Gaps 1;

QY 3546 GAAACGCACTTGGACGAGGAGGCTTCCGTTGAAGTCACGAGTCTGCTGACTC 3605
DB 3367 GAAAAGCCGAGGCTCTCAAGCCGATAGGAGCGTTCGATCTCATCTCGATTC 3426
QY 3606 CGGCGAGCTTGGAGCGCTTGGCGAACAACATCTTCTGCGCGCATGCTCAAGCT 3665
DB 3427 TGGCCCTGTGGACCATTCGAGAAACCTTATTTGCTGGGCTTATCTGACCTT 3486
QY 3666 CTTCGCGCGCTCAAGTTTCGAGCTTTCAAAGCGACGACTGCTCTGTGTAGCCA 3725
DB 3487 TTTGCTCGGCGTTCATGTCGCGCGAAGCGATGATGCTTCTTCGAGGGA---A 3543
QY 3726 TTACTCCGTTTCGACGCTAGCGCGCTTCAATGGCGCAAGCTTACAGACCAACTT 3785
DB 3544 TAAATTTGACTTGGACGCAAGACGCTTCGATGGGATCTTATAGCCGAAACAACTT 3603
QY 3786 GAAAGTCGAGGTCGAGAAATGTCGCGGATCATCGGACTCTCGTCTCGCTGAGAGGT 3845
DB 3604 GAAAGTCGAGGTCGAGGCGCGTGTGAGCTTTATAGGTTTATGTTTCCAAAGCGCGCT 3663
QY 3846 CTTCTTCGACCTCTCAGAGGCTCTCAAGATATTTGGGCGCTGCTACAGAGCAACT 3905
DB 3664 CACAGCTGATGTGTGCGCTCTAGCCACCGGACTTACGATGCAAGTTATATAAAGCTGAT 3723

QY 3906 CCTTA 3911
DB 3724 GATCTA 3729

RESULT 6

US-08-485-355B-47

Sequence 47, Application US/08485355B

Patent No. 6177075

GENERAL INFORMATION:

APPLICANT: Christlian, P. D., Gordon, K. H.J., Hanzlik, T. N.

TITLE OF INVENTION: Insect Viruses and Their Uses in Protecting Plants

NUMBER OF SEQUENCES: 57

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr Hobach Test Albritton & Herbert LLP

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/485,355B

FILING DATE: 07-Jun-1995

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/440,522

FILING DATE: 12-MAY-1995

APPLICATION NUMBER: US 08/089,372

FILING DATE: 08-JUL-1993

APPLICATION NUMBER: AU PL4081/92

FILING DATE: 14-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Treacartin, Richard F.

REGISTRATION NUMBER: 31,801

REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS

TELECOMMUNICATION INFORMATION:

TELEPHONE: (415) 781-1989

TELEFAX: (415) 398-3249

TELEX: 910 277299

INFORMATION FOR SEQ ID NO: 47:

SEQUENCE CHARACTERISTICS:

LENGTH: 2478 base pairs

TYPE: nucleic acid

STRANDEDNESS: unknown

TOPOLOGY: unknown

MOLECULE TYPE: DNA

FEATURE:

NAME/KEY: CDS

LOCATION: 283..753

SEQUENCE DESCRIPTION: SEQ ID NO: 47:

US-08-485-355B-47

Query Match 1.3%; Score 71.2; DB 3; Length 2478;
Best Local Similarity 78.7%; Pred. No. 6, 8e-08;
Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 5205 TAAAGCTCTGTGAGTCTCTCAAGCTTACTGTTAGTCTGCTGCGTTGATTCATTTC 5264
DB 2371 TAAAGCTCTGTGAGTCTCTCAAGCTTACTGTTAGTCTGCTGCGTTGATTCATTTC 2430
QY 5265 CCAAGCAAGAGGTCGCAACTAGTACGGGCGCCCTCGGATACCA 5312
DB 2431 CCAAGCGCAAGAGGTCGCAACTAGTACGGGCGCCCTCGGATACCA 2478

RESULT 7
US-08-485-355B-49

Sequence 49, Application US/08485355B
Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Christlan, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard P.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 2478 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 366..2306
SEQUENCE DESCRIPTION: SEQ ID NO: 49:
US-08-485-355B-49
Query Match 1.3%; Score 71.2; DB 3; Length 2478;
Best Local Similarity 78.7%; Pred. No. 6.8e-08;
Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 5205 TAAAGACTGTGGAGTCTCAACGTTACTCGTGTAGTCTGCGCGGTGGATTCATTTC 5264
DB 2371 TAAAGACTGTGGAGTCTCAACGTTACTCGTGTAGTCTGCGCGGTGGATTCATTTC 2430
QY 5265 CCAAGCAGCAAGAGGTGCGCAACTAGTACGCGCGCCCGCTGGATACCA 5312
DB 2431 CCAAGCAGCAAGAGGTGCGCAACTAGTACGCGCGCCCGCTGGATACCA 2478
RESULT 8
US-09-194-613-1
Sequence 1, Application US/09194613
Patent No. 6251654
GENERAL INFORMATION:
APPLICANT: GORDON, Karl H.
APPLICANT: HANZLIK, Terry N.
TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES
NUMBER OF SEQUENCES: 22

CORRESPONDENCE ADDRESS:
ADDRESSEE: McDermott, Will & Emery
STREET: 600 13th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3096
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194,613
FILING DATE: 30-NOV-1998
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Bucca Ph.D., Daniel
REGISTRATION NUMBER: 42,368
REFERENCE/DOCKET NUMBER: 50179-061
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-756-8000
TELEFAX: 202-756-8087
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2478 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-09-194-613-1

Query Match 1.3%; Score 71.2; DB 3; Length 2478;
Best Local Similarity 78.7%; Pred. No. 6.8e-08;
Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;
QY 5205 TAAAGACTGTGGAGTCTCAACGTTACTCGTGTAGTCTGCGCGGTGGATTCATTTC 5264
DB 2371 TAAAGACTGTGGAGTCTCAACGTTACTCGTGTAGTCTGCGCGGTGGATTCATTTC 2430
QY 5265 CCAAGCAGCAAGAGGTGCGCAACTAGTACGCGCGCCCGCTGGATACCA 5312
DB 2431 CCAAGCAGCAAGAGGTGCGCAACTAGTACGCGCGCCCGCTGGATACCA 2478
RESULT 9
US-08-485-355B-51
Sequence 51, Application US/08485355B
Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Christlan, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995


```
Db 346 CGGCGGCGTCAACCGTCCGAGCCACCGTCTCCACCGCCAGCGGTTTTCACGAGTT 405
Qy 1824 CACCTCTCTCCAGTCAAGAGGCTTCAGATGTAAGAGGCGCCGAGGCGGCAAGAC 1883
Db 406 CATCATCTCTCTCAACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 465
Qy 1884 GGGCACCCTTCATCGCGCGCTTCAGAGCGCGCGGCGGTAAGCACTTATGTCACCCAC 1943
Db 466 GGGTGAATCTGCGCGGATCGCGTGGAGCGCGGTAAGCACTTATGTCACCCAC 525
Qy 1944 CAGAGAACTGAGAGAGGCTATGAGCGGCGATCAACCGCGCTCCGCTCGGCTACG 2001
Db 526 GCGGAGACGCGCGGCGGCTCATGAACCGGTAAGAGACCGCTGGGCGCGCGCGCG 583

RESULT 12
US-09-252-991A-3828/c
; Sequence 3828, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252.991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074.788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094.190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 3828
; LENGTH: 567
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-3828

Query Match 0.98; Score 47.2; DB 4; Length 567;
Best Local Similarity 48.28; Pred. No. 0.032;
Matches 133; Conservative 0; Mismatches 143; Indels 0; Gaps 0;

Qy 1727 TCCCGAGTACCGGCGACCTTGAGGAGGATCTCAACACCGACGTCAGAGCTCAAGA 1786
Db 523 TCGCCGAGGCGATGAGCGCGGAGAGCCACCGTTCCGAGCGGCTTGACCAAGCTCGCG 464
Qy 1787 TCACCTTCGAGAAAGCGCCCTCAAGACATCGAGCGGCTCACCTCTCCCGAGTCAAGGCC 1846
Db 463 AGATTCGCAAGAGAGCGCTTCGAGAGAGATGGCGAAATCACCTCTCGCGATGGCAAGC 404
Qy 1847 TCGAGATGTAAGAGGCGCGCGCGAGCGGCAAGCGGACCGGACCTTATGCGCGCTTG 1906
Db 403 TGAAGGCTTCGAGCGGCTCCGCGGACCGCGGCAACCGGACCGGCAAGCGGATCG 344
Qy 1907 AGGCGCGGCGGTAAGCACTTATGAGGACCCACGAGGAGTGAAGAGGCTATGG 1966
Db 343 AGATTCGTCGCAAGCGTGAAGCAAGTTCGTTCCGCGCAAGGACCTACCGACCGCATCA 284
Qy 1967 ACCGCGGATCAACCGCGCTCCGCTCGGCTACGCG 2002
Db 283 ACCGCTGATAACGCGCGCGGATCCGCGACGAGATGC 248

RESULT 13
US-09-679-279-1
; Sequence 1, Application US/09679279
; Patent No. 6524841
; GENERAL INFORMATION:
; APPLICANT: McDaniel, Robert
; APPLICANT: Volchegursky, Yanina
; TITLE OF INVENTION: Recombinant Megalomicin Biosynthetic
; FILE REFERENCE: 300622004700
; CURRENT APPLICATION NUMBER: US/09/679.279

CURRENT FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/158.305
PRIOR FILING DATE: 1999-10-08
PRIOR APPLICATION NUMBER: US 60/190.024
PRIOR FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 34
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 47981
TYPE: DNA
ORGANISM: Micromonospora megalomicea
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(144)
OTHER INFORMATION: megBVI (megT), TDP-4-keto-6-deoxyglucose-2,3-dehydratase;
NAME/KEY: CDS
LOCATION: (144)...(2061)
OTHER INFORMATION: SEQ ID NO: 2= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (2061)...(2061)
OTHER INFORMATION: megDVI, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 3= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (2072)...(3382)
OTHER INFORMATION: megDI, rhodosaminyl transferase (eryCII homolog),
OTHER INFORMATION: TDP-megosamine glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 4= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (3462)...(4634)
OTHER INFORMATION: megQ (megY), mycarosyl acyltransferase, mycarose O-acetyltransferase;
OTHER INFORMATION: SEQ ID NO: 5= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (4651)...(5775)
OTHER INFORMATION: megDII, deoxy sugar transferase (eryCI, DnuI homolog),
OTHER INFORMATION: TDP-3-keto-6-deoxyhexose 3-aminotransaminase;
OTHER INFORMATION: SEQ ID NO: 6= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (5822)...(6595)
OTHER INFORMATION: megDIII, daunosaminyl-N,N-dimethyltransferase (eryCVI homolog);
OTHER INFORMATION: SEQ ID NO: 7= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (6592)...(7197)
OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 3,5-epimerase (eryBVI, dnuuI);
OTHER INFORMATION: SEQ ID NO: 8= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (7220)...(8206)
OTHER INFORMATION: megDV, TDP-hexose 4-ketoreductase (eryBIV, dnuv homolog),
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 9= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (8228)...(9220)
OTHER INFORMATION: megBII-1 (megDVI), TDP-4-keto-L-6-deoxy-hexose 2,3-reductase;
OTHER INFORMATION: SEQ ID NO: 10= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (9226)...(10479)
OTHER INFORMATION: megBV, mycarosyl transferase, mycarose glycosyltransferase;
OTHER INFORMATION: SEQ ID NO: 11= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (10483)...(11424)
OTHER INFORMATION: megBIV, TDP-hexose 4-ketoreductase;
OTHER INFORMATION: TDP-4-keto-6-deoxyhexose 4-ketoreductase;
OTHER INFORMATION: SEQ ID NO: 12= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (12181)...(22821)
OTHER INFORMATION: megAI, SEQ ID NO: 13= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (12505)...(13470)
OTHER INFORMATION: megAI, AT-L
NAME/KEY: misc feature
LOCATION: (13576)...(13791)
OTHER INFORMATION: megAI, ACP-L
NAME/KEY: misc feature
LOCATION: (13849)...(15126)
```

```

OTHER INFORMATION: megA1, KS1
NAME/KEY: misc feature
LOCATION: (15427)...(16476)
OTHER INFORMATION: megA1, AT1
NAME/KEY: misc feature
LOCATION: (17155)...(17694)
OTHER INFORMATION: megA1, KR1
NAME/KEY: misc feature
LOCATION: (17947)...(18207)
OTHER INFORMATION: megA1, ACP1
NAME/KEY: misc feature
LOCATION: (18268)...(19548)
OTHER INFORMATION: megA1, KS2
NAME/KEY: misc feature
LOCATION: (19876)...(20910)
OTHER INFORMATION: megA1, AT2
NAME/KEY: misc feature
LOCATION: (21517)...(22053)
OTHER INFORMATION: megA1, KR2
NAME/KEY: misc feature
LOCATION: (22318)...(22575)
OTHER INFORMATION: megA1, ACP2
NAME/KEY: CDS
LOCATION: (22867)...(33555)
OTHER INFORMATION: megA1; SEQ ID NO: 14= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (22957)...(24237)
OTHER INFORMATION: megA1, KS3
NAME/KEY: misc feature
LOCATION: (24574)...(25581)
OTHER INFORMATION: megA1, AT3
NAME/KEY: misc feature
LOCATION: (26230)...(26733)
OTHER INFORMATION: megA1, KR3 (inactive)
NAME/KEY: misc feature
LOCATION: (26958)...(27258)
OTHER INFORMATION: megA1, ACP3
NAME/KEY: misc feature
LOCATION: (27393)...(28590)
OTHER INFORMATION: megA1, KS4
NAME/KEY: misc feature
LOCATION: (28857)...(29931)
OTHER INFORMATION: megA1, AT4
NAME/KEY: misc feature
LOCATION: (29953)...(30477)
OTHER INFORMATION: megA1, DH4
NAME/KEY: misc feature
LOCATION: (31356)...(32244)
OTHER INFORMATION: megA1, ER4
NAME/KEY: misc feature
LOCATION: (32257)...(32799)
OTHER INFORMATION: megA1, KR4
NAME/KEY: misc feature
LOCATION: (33052)...(33312)
OTHER INFORMATION: megA1, ACP4
NAME/KEY: CDS
LOCATION: (33666)...(43271)
OTHER INFORMATION: megA1; SEQ ID NO: 15= translated amino acid sequence
NAME/KEY: misc feature
LOCATION: (33780)...(35027)
OTHER INFORMATION: megA1, KS5
NAME/KEY: misc feature
LOCATION: (35385)...(36419)
OTHER INFORMATION: megA1, AT5
NAME/KEY: misc feature
LOCATION: (37068)...(37604)
OTHER INFORMATION: megA1, KR5
NAME/KEY: misc feature
LOCATION: (37860)...(38120)
OTHER INFORMATION: megA1, ACP5
NAME/KEY: misc feature
LOCATION: (38187)...(39470)
OTHER INFORMATION: megA1, KS6

```

```

NAME/KEY: misc feature
LOCATION: (39755)...(40811)
OTHER INFORMATION: megA1, AT6
NAME/KEY: misc feature
LOCATION: (41406)...(41936)
OTHER INFORMATION: megA1, KR6
NAME/KEY: misc feature
LOCATION: (42168)...(42425)
OTHER INFORMATION: megA1, ACP6
NAME/KEY: misc feature
LOCATION: (42585)...(43271)
OTHER INFORMATION: megA1, TR
NAME/KEY: CDS
LOCATION: (43268)...(44344)
OTHER INFORMATION: megC11, TDP-4-keto-6-deoxyglucose 3,4-isomerase;
OTHER INFORMATION: SEQ ID NO: 16= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (44355)...(45623)
OTHER INFORMATION: SEQ ID NO: 17= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (45620)...(46591)
OTHER INFORMATION: megB11-2(megB11), TDP-4-keto-6-deoxy-L-glucose 2,3 dehydratase,
OTHER INFORMATION: TDP-4-keto-6-deoxyglucose 2,3 dehydratase;
OTHER INFORMATION: SEQ ID NO: 18= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (46660)...(47403)
OTHER INFORMATION: megH, TE11; SEQ ID NO: 19= translated amino acid sequence
NAME/KEY: CDS
LOCATION: (47411)...(47980)
OTHER INFORMATION: megF, C-6 hydroxylase; SEQ ID NO: 20= translated amino acid seq
US-09-679-279-1

```

```

Query Match 0.9%; Score 47.2; DB 4; Length 47981;
Best Local Similarity 46.2%; Pred. No. 0.16;
Matches 157; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

```

```

QY 1651 CCGGTAACCTTCACCTTCCTCTGTGAGAGAGAGCCCAAGTTTGAGCGCCGACCTTGTGGAC 1710
DB 30391 CCGGCGCCACCGCCCTTACGTTGACCGACCACTAGACACCTTGGCGAAGCTGGCTTAC 30450
QY 1711 GCGACCGCGGCGCTCACTCCCGAGTACCGCGCCACTTTCAGAGAGGTCTCAACCGAC 1770
DB 30451 GAGTACGGGCGCGGCTTCCAGGCGCTGCGCGCGGTGACAGACGCGACCTGCTTAC 30510
QY 1771 GTCAAGCAGCTCAAGATCACTCTGAGAAAGCGCTCAAGACCATGACGAGCTCAACCTTC 1830
DB 30511 GCGAGGTGTCTCCGTGAGCGCGTCCGAGGAGGAGGTACGCTTGAACCGGTGCTCGAC 30570
QY 1831 TCCCAAGTACAGAGGCTCTGAGATGATACAGAGGCGCCGACGAGCGCAAGCGGACCC 1890
DB 30571 GCGGTGCGCCAGACCTTCGCGCTGACCACTCCGCGCGGAGAGCTTCCCTTGGCTGCG 30630
QY 1891 CTCATCGCGCGCTTGAAGCGCGCGGCGGTAAAGCACTTACGTGGACCCACAGAGAA 1950
DB 30631 CGGGGCTACCTTGCACCGACCGGCGGCTGCGGTACGAGTGTGGAGACCCCGCC 30690
QY 1951 CTGAGAGAGCTTATGAGACCGCGGATCAACCGCGCTCG 1990
DB 30691 GGACCGGACGCGGTGCGCGGTCAACCGACCGACCG 30730

```

```

RESULT 14
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FIEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TUBERCULOSIS

```

FILE REFERENCE: 24366-20007.00
 CURRENT APPLICATION NUMBER: US/09/103,840A
 CURRENT FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 2
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2
 LENGTH: 4403765
 TYPE: DNA
 ORGANISM: Mycobacterium tuberculosis
 FEATURE:
 OTHER INFORMATION: CDC 1551
 OTHER INFORMATION: "n" bases at various positions throughout the sequence
 OTHER INFORMATION: represent a, t, c or g
 US-09-103-840A-2

Query Match 0.9%; Score 46.8; DB 3; Length 4403765;
 Best Local Similarity 50.4%; Pred. No. 1.1;
 Matches 114; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 345 CGCCCCCGTGCAGACTATACGGGTCACCAAGTACGGACCCCGGACGGCTCGGACA 404
 DB 3922657 CGACGCGTGCAGGATACGGCGCTTGCCTCCGACCGCGCCACCTGGCCG 3922598
 QY 405 CATTACGGCCTTAGAGTCTAGTCCGTCGACAGCGCGCCGAGTTCAAGGCCGACGC 464
 DB 3922597 CATTGCGCGCTGACCGCGCCCGCGCTTACCGCGCCCGCGCATTTGCGCTG 3922538
 QY 465 CTCAGTCTGCGCAAGGATTCCTCCGACCTTCTGCGTGCAGAGTGCAGTCTTGG 524
 DB 3922537 CGCGCGCGGCGGACCGTCCCGCGCTTCCACCTTTGGCGCGGTGCCCCGCAAGCT 3922478
 QY 525 CGCGTTCAATCGCGCGCTTGAATGCGCAATCACTCCCTTATGAC 570
 DB 3922477 CTGTGCGAGTGCAGCGCGGTATGCGGTGCGCGCGCCCGCAATGCC 3922432

RESULT 15
 US-09-372-422A-23
 Sequence 23, Application US/09372422A
 Patent No. 6313375
 GENERAL INFORMATION:
 APPLICANT: Rudolf Jung
 TITLE OF INVENTION: Maize Aquaporins and Uses Thereof
 FILE REFERENCE: 0919
 CURRENT APPLICATION NUMBER: US/09/372,422A
 CURRENT FILING DATE: 1999-08-11
 PRIOR APPLICATION NUMBER: US 60/098,692
 PRIOR FILING DATE: 1998-08-31
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 23
 LENGTH: 1193
 TYPE: DNA
 ORGANISM: Zea mays
 FEATURE:
 NAME/KEY: CDS
 LOCATION: (88)...(838)
 US-09-372-422A-23

Query Match 0.9%; Score 45.8; DB 4; Length 1193;
 Best Local Similarity 47.1%; Pred. No. 0.093;
 Matches 140; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 3775 ACCAATGATTGAGTGCAGAGAAATCGTACATCGGATCGCTCTGCTCC 3834
 DB 496 ACCACCGAGCTGGAGGCGTGTGATCGATCTCACTTCGCGCTGTCTACACC 555
 QY 3835 GCTGAGAGAGTGTCTGACCTGTGACAGAGCGCTCTCAAGATATTGGCGCTGTAC 3894
 DB 556 GTGTAAGCCACCGCGCGCCAGAACAGAGGCTCTCTCGGACCATCGCGCATGCC 615
 QY 3895 ACAAGCAACTCTTTACTCAAGTACGTGAGGCTGTGAGAGCATCACCAAGGCTCG 3954

DB 616 ATCGGCTTATGATGCGAGGCGCAACATCTCGCGCGGAGGCGCTTACAGCGGCTCATG 675
 QY 3955 AGTACGCGCGCTTACCAAGGCTCTGTGCAATATGACATGCTACTACATTAACGGC 4014
 DB 676 AACCGCGCGCTCTTGGGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 735
 QY 4015 CCGAGTCTGCGGCGGCTTACATCAATCAAGCTGTGTGCTTTGGGCGGCGGCGGCTTC 4071
 DB 736 TACTGAGTGCAGCGCGCTCATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 792

Search completed: January 15, 2004, 13:22:23
 Job time : 257 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 15, 2004, 11:23:16 ; Search time 1197 Seconds
(without alignments)
15642.057 Million cell updates/sec

Title: US-09-991-262-39

Perfect score: 5312
Sequence: 1 GTTGTGCTCCCGGACG.....CGGGCCCTTGAGTACCA 5312

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapept 1.0

Searched: 2324096 seqs, 1762381658 residues

Total number of hits satisfying chosen parameters: 4648192

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PTCT_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PTCT_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 18: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	5312	100.0	5312	11	US-09-991-262-39 Sequence 39, Appl
2	5312	100.0	5312	11	US-09-991-262-41 Sequence 41, Appl
3	5312	100.0	5312	11	US-09-991-262-43 Sequence 43, Appl
4	5206	98.0	5368	11	US-09-991-262-45 Sequence 45, Appl
5	71.2	1.3	2478	11	US-09-991-262-47 Sequence 47, Appl
6	71.2	1.3	2478	11	US-09-991-262-49 Sequence 49, Appl
7	71.2	1.3	2479	11	US-09-991-262-51 Sequence 51, Appl
8	51.6	1.0	627	15	US-10-156-761-3243 Sequence 3243, Ap
9	51.6	1.0	9025608	15	US-10-156-761-1 Sequence 1, Appl
10	51.2	1.0	1650	15	US-10-156-761-6555 Sequence 6555, Ap
11	49.8	0.9	918	12	US-10-369-493-42399 Sequence 42399, A
12	49.8	0.9	1449	15	US-10-156-761-4793 Sequence 4793, A
13	49.8	0.9	9025608	15	US-10-156-761-1 Sequence 1, Appl
14	49.6	0.9	1467	12	US-10-369-493-43237 Sequence 43237, A
15	49	0.9	1047	15	US-10-156-761-4716 Sequence 4716, Ap

Result No.	Score	Query Match	Length	DB ID	Description
16	47.8	0.9	1560	15	US-10-156-761-5120 Sequence 5120, Ap
17	47.6	0.9	1005	15	US-10-156-761-6273 Sequence 6273, Ap
18	47.6	0.9	1170	13	US-10-409-701-16 Sequence 16, Appl
19	46.4	0.9	1049	13	US-10-140-472-358 Sequence 358, Appl
20	46.4	0.9	1049	13	US-10-141-761-358 Sequence 358, Appl
21	46.4	0.9	1049	13	US-10-142-885-358 Sequence 358, Appl
22	46.4	0.9	1049	13	US-10-158-790-358 Sequence 358, Appl
23	46.4	0.9	1049	13	US-10-137-871-358 Sequence 358, Appl
24	46.4	0.9	1049	13	US-10-140-805-358 Sequence 358, Appl
25	46.4	0.9	1049	13	US-10-140-864-358 Sequence 358, Appl
26	46.4	0.9	1049	13	US-10-140-923-358 Sequence 358, Appl
27	46.4	0.9	1049	13	US-10-141-756-358 Sequence 358, Appl
28	46.4	0.9	1049	13	US-10-141-759-358 Sequence 358, Appl
29	46.4	0.9	1049	15	US-10-123-155-358 Sequence 358, Appl
30	46.4	0.9	1049	16	US-10-146-731-358 Sequence 358, Appl
31	46.2	0.9	870	15	US-10-156-761-3286 Sequence 3286, Ap
32	45.8	0.9	768	12	US-10-369-493-32055 Sequence 32055, A
33	45.6	0.9	1563	15	US-10-156-761-3340 Sequence 3340, Ap
34	45.6	0.9	2924	12	US-10-310-154-149 Sequence 149, Ap
35	45	0.8	714	15	US-10-156-761-6500 Sequence 6500, Ap
36	44.6	0.8	520	15	US-10-184-644-332 Sequence 332, Ap
37	44.6	0.8	520	15	US-10-184-634-332 Sequence 332, Ap
38	44	0.8	3138	15	US-10-156-761-2538 Sequence 2538, Ap
39	44	0.8	77536	13	US-09-940-3168-1 Sequence 1, Appl
40	43.8	0.8	1062	12	US-10-369-493-42337 Sequence 42337, A
41	43.6	0.8	1323	15	US-10-156-761-6130 Sequence 6130, Ap
42	43.6	0.8	1446	15	US-10-156-761-3406 Sequence 3406, Ap
43	43.6	0.8	1770	12	US-10-369-493-31941 Sequence 31941, A
44	43.4	0.8	603	15	US-10-156-761-3254 Sequence 3254, Ap
45	43.4	0.8	1287	15	US-10-156-761-2905 Sequence 2905, Ap

ALIGNMENTS

RESULT 1
US-09-991-262-39
Sequence 39, Application US/09991262
Publicat No. US20030041349A1
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H. J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr Hobbach Teet Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/991,262
FILING DATE: 20-NO. US20030041349A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/234,238
FILING DATE: 20-JAN-1999
APPLICATION NUMBER: US 08/485,355
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.

REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-58631-3/RFT/DSS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEEX: 910 272299
 INFORMATION FOR SEQ ID NO: 39:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 5312 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: unknown
 TOPOLOGY: unknown
 MOLECULE TYPE: DNA
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 37..5148
 SEQUENCE DESCRIPTION: SEQ ID NO: 39:
 US-09-991-262-39

Query Match 100.0%; Score 5312; DB 11; Length 5312;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 5312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTGCTCTCCCGGCGGTAATATAGGGGAAACAATGTACGGGAAAGGACGACAGTGC 60
 DB 1 GTTCTGCTCTCCCGGCGGTAATATAGGGGAAACAATGTACGGGAAAGGACGACAGTGC 60
 QY 61 GCGGCTGTCTACGCGCGGCGGAGATGTGCGCTACGCGAAGTACTGCGACGAGAGAGTGC 120
 DB 61 GCGGCTGTCTACGCGCGGCGGAGATGTGCGCTACGCGAAGTACTGCGACGAGAGAGTGC 120
 QY 121 AAGTTGACTTCCGCGCGCACTGAAAGCACTAGAAAACCTCCACAGACTGTACTATCCG 180
 DB 121 AAGTTGACTTCCGCGCGCACTGAAAGCACTAGAAAACCTCCACAGACTGTACTATCCG 180
 QY 181 CTGCGCTTCAAAAGGGGGCACTTTACCCCGGACACAAACCCGATCTCGGCGGGACCAA 240
 DB 181 CTGCGCTTCAAAAGGGGGCACTTTACCCCGGACACAAACCCGATCTCGGCGGGACCAA 240
 QY 241 CGGTGCGAGAGAGTTCTGCAACAATTTGCGCGAGGGGAGGTAGACAGTGTGAGATA 300
 DB 241 CGGTGCGAGAGAGTTCTGCAACAATTTGCGCGAGGGGAGGTAGACAGTGTGAGATA 300
 QY 301 GGGCGGTCTCTGCAACGCGCACTTAAGTACATGAGGAGACCGGACCGCGCTGCGACAC 360
 DB 301 GGGCGGTCTCTGCAACGCGCACTTAAGTACATGAGGAGACCGGACCGCGCTGCGACAC 360
 QY 361 TATCAGCGGTGCAACAGTACGCGACCGCGGAGCTGCGACACATTAAGGCTTTAGAG 420
 DB 361 TATCAGCGGTGCAACAGTACGCGACCGCGGAGCTGCGACACATTAAGGCTTTAGAG 420
 QY 421 TCTGATCCGCTGCGACAGCGCGGCGCGAGTTCAAGGCGGAGCTCACTGCTGCGCAAC 480
 DB 421 TCTGATCCGCTGCGACAGCGCGGCGCGAGTTCAAGGCGGAGCTCACTGCTGCGCAAC 480
 QY 481 GGCATTGCTCCCGCACTTCTGCGTGAAGGAGTCCGCTCTTGGCGTTAAATTCGGGC 540
 DB 481 GGCATTGCTCCCGCACTTCTGCGTGAAGGAGTCCGCTCTTGGCGTTAAATTCGGGC 540
 QY 541 GTTGAATGCGCAATCACTCCCTTATGACGTGACCTTGAAGAGCTGGCCATGCGTTT 600
 DB 541 GTTGAATGCGCAATCACTCCCTTATGACGTGACCTTGAAGAGCTGGCCATGCGTTT 600
 QY 601 GAGAACACGAGCTTCAATGCTCGCGGCTTCAATGCAATGCGAAGAGAGCTGCTAC 660
 DB 601 GAGAACACGAGCTTCAATGCTCGCGGCTTCAATGCAATGCGAAGAGAGCTGCTAC 660
 QY 661 ATGACAAAGTGTAAATGCGAGCTCGGCTTCAACGCTTCAAGTAAATGAGAGCTATG 720
 DB 661 ATGACAAAGTGTAAATGCGAGCTCGGCTTCAACGCTTCAAGTAAATGAGAGCTATG 720
 QY 721 GCTGTGAAGAGCTCGGCTTCAAGGAGGAGAGCTCGCTTCACTTCCGAGTTGAGAC 780
 DB 721 GCTGTGAAGAGCTCGGCTTCAAGGAGGAGAGCTCGCTTCACTTCCGAGTTGAGAC 780

DB 721 GCTGTGAAGAGCTCGGCTTCAAGGAGGAGAGCTCGCTTCACTTCCGAGTTGAGAC 780
 QY 781 TTGATCAACGAGAGCCAAAGCGGCGCATCGAAGGCTGGCGCGCGGCTCTACTCC 840
 DB 781 TTGATCAACGAGAGCCAAAGCGGCGCATCGAAGGCTGGCGCGCGGCTCTACTCC 840
 QY 841 AGACGGCGGCTGATTTTCTCCGCGGAGAGAGCTGAGGAGTGAAGGCTTACAGAGCTTC 900
 DB 841 AGACGGCGGCTGATTTTCTCCGCGGAGAGAGCTGAGGAGTGAAGGCTTACAGAGCTTC 900
 QY 901 CACACATGAGCTGCTACTACTGATGAGAGAACTACCCCACTCGTTGGTTCTACTC 960
 DB 901 CACACATGAGCTGCTACTACTGATGAGAGAACTACCCCACTCGTTGGTTCTACTC 960
 QY 961 CATATGAAGATCCAGAGGCGGCGCATCGAGATTTGAGCTGCGATCACTCGCGGCGCA 1020
 DB 961 CATATGAAGATCCAGAGGCGGCGCATCGAGATTTGAGCTGCGATCACTCGCGGCGCA 1020
 QY 1021 CTTGAGAGACCGCATGCTGGCGCGTCTGCCAAGAGAGCTCCCAAGGCTCTGCAAAATCCA 1080
 DB 1021 CTTGAGAGACCGCATGCTGGCGCGTCTGCCAAGAGAGCTCCCAAGGCTCTGCAAAATCCA 1080
 QY 1081 AACATCTTTTATTAAGCGCGAGCGGCTGAGGACCTAAGACATCCCTTACGTACAG 1140
 DB 1081 AACATCTTTTATTAAGCGCGAGCGGCTGAGGACCTAAGACATCCCTTACGTACAG 1140
 QY 1141 CACAAAGTCAACATGCTGCTCAATTTTATGCAACCGGCTCTGAGAGAGAACTAGTGCAC 1200
 DB 1141 CACAAAGTCAACATGCTGCTCAATTTTATGCAACCGGCTCTGAGAGAGAACTAGTGCAC 1200
 QY 1201 ATGACGCTCTTAAATGTCGTCGCGCGCGTGAAGGCTGCGGCAATCGTGTGCTCAGAA 1260
 DB 1201 ATGACGCTCTTAAATGTCGTCGCGCGCGTGAAGGCTGCGGCAATCGTGTGCTCAGAA 1260
 QY 1261 GTACCGAGAGAGCTCTGGAACATCTGACCGGCTGACCTGATCCGACATGTCGTCTT 1320
 DB 1261 GTACCGAGAGAGCTCTGGAACATCTGACCGGCTGACCTGATCCGACATGTCGTCTT 1320
 QY 1321 TAGGCTCTCAACATCAATCGAGCGCGAGGCTGCGGCTGCTGCAAGACCGCAAGAC 1380
 DB 1321 TAGGCTCTCAACATCAATCGAGCGCGAGGCTGCGGCTGCTGCAAGACCGCAAGAC 1380
 QY 1381 GACGCTCTTGAAGAGACTTGTCTGAGAGAGTCTCAAGAGCTCTTGGGCTCTGTTC 1440
 DB 1381 GACGCTCTTGAAGAGACTTGTCTGAGAGAGTCTCAAGAGCTCTTGGGCTCTGTTC 1440
 QY 1441 GGTCTGCGCAACTCAAGGCGCGAGCTGCTCTTAAAGGCGGCGTGAAGTAAAGTAC 1500
 DB 1441 GGTCTGCGCAACTCAAGGCGCGAGCTGCTCTTAAAGGCGGCGTGAAGTAAAGTAC 1500
 QY 1501 CGAGTCACTCGCTCGAGAGCATTAATCTGCGAGCTGCTGCTCCCTGAACAGTTCGAC 1560
 DB 1501 CGAGTCACTCGCTCGAGAGCATTAATCTGCGAGCTGCTGCTCCCTGAACAGTTCGAC 1560
 QY 1561 TTCCTGCGTCCGCGGTACCACTGCGCGGCTTTTCAAGACAGGAAAGCTTGAAGTTC 1620
 DB 1561 TTCCTGCGTCCGCGGTACCACTGCGCGGCTTTTCAAGACAGGAAAGCTTGAAGTTC 1620
 QY 1621 CTTGCGAAGCTGCTGCTGCAAGAGAGTCCGCTTCACTGCTGCTGAGAGAGAG 1680
 DB 1621 CTTGCGAAGCTGCTGCTGCAAGAGAGTCCGCTTCACTGCTGCTGAGAGAGAG 1680
 QY 1681 CCCCAAGTTCGACGCGCGACTTGTGCGACGCGACCGGCTCACTCCCGAGTACCGC 1740
 DB 1681 CCCCAAGTTCGACGCGCGACTTGTGCGACGCGACCGGCTCACTCCCGAGTACCGC 1740
 QY 1741 GCGACTTGAAGGAGGCTTCAACACCGAGCTCAAGAGAGCTCAAGATCACTCTGAGAAC 1800
 DB 1741 GCGACTTGAAGGAGGCTTCAACACCGAGCTCAAGAGAGCTCAAGATCACTCTGAGAAC 1800
 QY 1801 GCGCTCAAGACATCGAGCGGCGTCACTGCTCCCGAGTCAAGAGGCTCGAGAGTACGAG 1860
 DB 1801 GCGCTCAAGACATCGAGCGGCGTCACTGCTCCCGAGTCAAGAGGCTCGAGAGTACGAG 1860

QY	1861	GGCCCCGACGACAGCGGCGAAGA	CGGGGCAACCTCTATGCCCGCCCTTTAGAGGCGCGGGGGGT	1320
Dp	1861	GGCCCCGACGACAGCGGCGAAGACGGGAC	CCTCTATCGCGCCCTTTAGAGGCGCGGGGGGT	1320
QY	1921	AAAGCATTTCACGTGAGCACCCACGACGAGAAC	CTGAGAGAGGCTATGGAACCGGCGGATCAAA	1980
Dp	1921	AAAGCATTTCACGTGAGCACCCACGAGAAC	CTGAGAGAGGCTATGGAACCGGCGGATCAAA	1980
QY	1981	CCGCGCTCCGCTTCGGCTACGCAACATGTGCGC	CTTGCGATTCTCCGTCTGTCCACCGCC	2040
Dp	1981	CCGCGCTCCGCTTCGGCTACGCAACATGTGCGC	CTTGCGATTCTCCGTCTGTGTCCACCGCC	2040
QY	2041	GAGGGCGCCCCCTTTGCGCTACCGTGGATTATGCA	CGAGAGGCTTCATGTTCCCGCTGTGTAC	2100
Dp	2041	GAGGGCGCCCCCTTTGCGCTACCGTGGATTATGCA	CGAGAGGCTTCATGTTCCCGCTGTGTAC	2100
QY	2101	GTGCGCATCGTGACGCGCTTGTCCCGAGCTCAC	GAAATAGTCTTTGTAGGGGACGTGCAC	2160
Dp	2101	GTGCGCATCGTGACGCGCTTGTCCCGAGCTCAC	GAAATAGTCTTTGTAGGGGACGTGCAC	2160
QY	2161	CAAAATCGGGTTTATAGACTTCGCAAGGCA	CAAGCGCGGAACTACGCGCTGTGTGGAAGTGC	2220
Dp	2161	CAAAATCGGGTTTATAGACTTCGCAAGGCA	CAAGCGCGGAACTACGCGCTGTGTGGAAGTGC	2220
QY	2221	GTTTAAAGCAGTGC	CGGTCCGCGCATCTTCAACCAAA	2280
Dp	2221	GTTTAAAGCAGTGC	CGGTCCGCGCATCTTCAACCAAA	2280
QY	2281	GGCACACACGTTTTTTCAGAGCTTGTAC	CCCGGGGTGCACACCACTTACGGGTGCGTGC	2340
Dp	2281	GGCACACACGTTTTTTCAGAGCTTGTAC	CCCGGGGTGCACACCACTTACGGGTGCGTGC	2340
QY	2341	TCCATTCAGCCACGTCGCGCCCGACGAC	CTACCGGACGACGAGGCGGCAAA	2400
Dp	2341	TCCATTCAGCCACGTCGCGCCCGACGAC	CTACCGGACGACGAGGCGGCAAA	2400
QY	2401	CAGAGGAAAAAGTGC	CGCGCACGCGGCTGAGGGCTGAGTGTGCAC	2460
Dp	2401	CAGAGGAAAAAGTGC	CGCGCACGCGGCTGAGGGCTGAGTGTGCAC	2460
QY	2461	CGCATTTTGGGTCTGTGCATCTCTGCA	TTATCAACA	2520
Dp	2461	CGCATTTTGGGTCTGTGCATCTCTGCA	TTATCAACA	2520
QY	2521	GCTAGAGAAAGTGC	GACCTTCTAGTGGCGCATCA	2580
Dp	2521	GCTAGAGAAAGTGC	GACCTTCTAGTGGCGCATCA	2580
QY	2581	GACCCGACAGGTGACATTTGAGAGAC	CACTCAACCAATACGCGGAAAGCCGAGGTGTTTACA	2640
Dp	2581	GACCCGACAGGTGACATTTGAGAGAC	CACTCAACCAATACGCGGAAAGCCGAGGTGTTTACA	2640
QY	2641	GACATCCCTTGACCCCTTGAGATCA	CGACTGTCAAA	2700
Dp	2641	GACATCCCTTGACCCCTTGAGATCA	CGACTGTCAAA	2700
QY	2701	GAAGTGATGGCAACGATAC	CCCCCGCAGAGTGCAC	2760
Dp	2701	GAAGTGATGGCAACGATAC	CCCCCGCAGAGTGCAC	2760
QY	2761	CGCAAGAACTTCGGGGACCAAC	CCGACGTGTGCTCTTTGGCGAAGACCGGCTAC	2820
Dp	2761	CGCAAGAACTTCGGGGACCAAC	CCGACGTGTGCTCTTTGGCGAAGACCGGCTAC	2820
QY	2821	GAGGTGTTGGCGGTCTGTGC	CAAAATCAACGTAAGGCTTTGCGCAAC	2880
Dp	2821	GAGGTGTTGGCGGTCTGTGC	CAAAATCAACGTAAGGCTTTGCGCAAC	2880
QY	2881	AAAGCCGACATTAAGGCGCTTCCAGAGAGGGGT	ACATGAGGCTTCAACACGCTGTAC	2940
Dp	2881	AAAGCCGACATTAAGGCGCTTCCAGAGAGGGGT	ACATGAGGCTTCAACACGCTGTAC	2940

QY	2241	AAAACACGAGGCGCTCCAGAGCGCTGTTGTCCTCCGCTACACCAAGCGAAGCGCTGACCTGCGC	3000
Db	2241	AAAACACGAGGCGCTCCAGAGCGCTGTTGTCCTCCGCTACACCAAGCGAAGCGCTGACCTGCGC	3000
QY	3001	CTACACGAGCTAAGAGAGAGCTGCAACGATGTCTAACTCGCTTGGACGACATTTGGGAC	3060
Db	3001	CTACACGAGCTAAGAGAGAGCTGCAACGATGTCTAACTCGCTTGGACGACATTTGGGAC	3060
QY	3061	TGGACTGTCACTGAMAAGCGCCCGGTGACCGAGCTGTCTTGGAGCCGAGCTTCAAGTTCAAC	3120
Db	3061	TGGACTGTCACTGAMAAGCGCCCGGTGACCGAGCTGTCTTGGAGCCGAGCTTCAAGTTCAAC	3120
QY	3121	CAAGCGCGCGGCACCGCTTGAAGAGCTGTGTGAGGCAACGACCCCTTCACTCCGTGACATA	3180
Db	3121	CAAGCGCGCGGCACCGCTTGAAGAGCTGTGTGAGGCAACGACCCCTTCACTCCGTGACATA	3180
QY	3181	GACTTCCCTTAATGAAGACTCAGCAGAAAGTGTCGCGCCAGCGGATCAATTAAGGGGAAAGT	3240
Db	3181	GACTTCCCTTAATGAAGACTCAGCAGAAAGTGTCGCGCCAGCGGATCAATTAAGGGGAAAGT	3240
QY	3241	GGGACGAGGAGATCGCGGCTCACTCAAGTCTTCTCAACTTGGTCTGCGCGCTTGGATAAGC	3300
Db	3241	GGGACGAGGAGATCGCGGCTCACTCAAGTCTTCTCAACTTGGTCTGCGCGCTTGGATAAGC	3300
QY	3301	ATACTCCGAGAGATATCTCCGTATCCGGAGCGCGACCGGTCCGGTACAGCAACGCTCTCCC	3360
Db	3301	ATACTCCGAGAGATATCTCCGTATCCGGAGCGCGACCGGTCCGGTACAGCAACGCTCTCCC	3360
QY	3361	GACGAGAAGAGGCGATGCTGTCTGAAAGCGAAGATCAATCAAGTCCCAACAGCGCAAGTTC	3420
Db	3361	GACGAGAAGAGGCGATGCTGTCTGAAAGCGAAGATCAATCAAGTCCCAACAGCGCAAGTTC	3420
QY	3421	GTCTCCGCGGAGCTGGAACCGAGTTTGGACAACCGCCCAATAACAGAGTGAGCTGCTTTC	3480
Db	3421	GTCTCCGCGGAGCTGGAACCGAGTTTGGACAACCGCCCAATAACAGAGTGAGCTGCTTTC	3480
QY	3481	GCCGCGCTTTTGAAGCGCATGTGGACCGCTGACAGCTGCGCTTAATCTATTCAGAGAACGG	3540
Db	3481	GCCGCGCTTTTGAAGCGCATGTGGACCGCTGACAGCTGCGCTTAATCTATTCAGAGAACGG	3540
QY	3541	TGTGGGAAACGCACTTGCGAGCGAAGGATGTAGAGTCCGTTGAAGTGAAGCGTCTGCTC	3600
Db	3541	TGTGGGAAACGCACTTGCGAGCGAAGGATGTAGAGTCCGTTGAAGTGAAGCGTCTGCTC	3600
QY	3601	GACTCCGCGCGAGCTTGGAGCGCTTGGCGCAACACATCTTCTGCGCGCGGTATGCTC	3660
Db	3601	GACTCCGCGCGAGCTTGGAGCGCTTGGCGCAACACATCTTCTGCGCGCGGTATGCTC	3660
QY	3661	ACGCTCTTCCGCGGCGGTCAAGTTGCGAGCTTTCAAAGGCGACGACTCGCTCTCTGTGAGT	3720
Db	3661	ACGCTCTTCCGCGGCGGTCAAGTTGCGAGCTTTCAAAGGCGACGACTCGCTCTCTGTGAGT	3720
QY	3721	AGCCATTACCTCCGTTTTCGAGCTGAGCCGCTTTCACATGGGCGAAAGTTACAAAGCAA	3780
Db	3721	AGCCATTACCTCCGTTTTCGAGCTGAGCCGCTTTCACATGGGCGAAAGTTACAAAGCAA	3780
QY	3781	CATTGGAAGGTGAGAGTGCAGAAATATGATGCGGTGACATCGGACTCTCTGCTCGGTAG	3840
Db	3781	CATTGGAAGGTGAGAGTGCAGAAATATGATGCGGTGACATCGGACTCTCTGCTCGGTAG	3840
QY	3841	CAGGTGCTCTCGAGCCGCTGAGAGCGCTTCAAGATATTTGGGCGCTGTGTAACAAGC	3900
Db	3841	CAGGTGCTCTCGAGCCGCTGAGAGCGCTTCAAGATATTTGGGCGCTGTGTAACAAGC	3900
QY	3901	GAATCTCTTATCTCCAAGTATGTGAGAGGCTGTGAGAGACATCACCAAGGCTGTGAGTAC	3960
Db	3901	GAATCTCTTATCTCCAAGTATGTGAGAGGCTGTGAGAGACATCACCAAGGCTGTGAGTAC	3960
QY	3961	GCCGCGTACACAGCGCTCTGTGACATGTGACAGATGTCTACTCAATTAACGCGCGGAG	4020
Db	3961	GCCGCGTACACAGCGCTCTGTGACATGTGACAGATGTCTACTCAATTAACGCGCGGAG	4020
QY	4021	TCTGGGCGTACATCATGACGCTGTGTTGCTTTGGGCGCGGACATTCCTCGTTGAA	4080

```

Db      4021 TCTGCGCGCTACATCATCAAGCTGTTGTTGGGCGCGCACTTCCGTTTGA 4080
Qy      4081 CAATCGCGGTGGTGGTGGTCCCATGTGCAAGGACCGGACGCTTACAGACGATATCG 4140
Db      4081 CAATCGCGGTGGTGGTGGTCCCATGTGCAAGGACCGGACGCTTACAGACGATATCG 4140
Qy      4141 GCTAACGTGGCGCATGTGCTTGAACCAAGTCTTGAAGCCCGCGCAAGCGCGCGCG 4200
Db      4141 GCTAACGTGGCGCATGTGCTTGAACCAAGTCTTGAAGCCCGCGCAAGCGCGCGCG 4200
Qy      4201 GCGAGTTTGGTGGCAATGTGGGAAGCCGGAAACGCTTCTTCACTTACCGGAAAGCT 4260
Db      4201 GCGAGTTTGGTGGCAATGTGGGAAGCCGGAAACGCTTCTTCACTTACCGGAAAGCT 4260
Qy      4261 GGTGTTTCTGCACTCAAGCCCAAGTGGCACTGCGCCCGGAGTCTCCATGG 4320
Db      4261 GGTGTTTCTGCACTCAAGCCCAAGTGGCACTGCGCCCGGAGTCTCCATGG 4320
Qy      4321 GATGCACTTGAGCCCAAGCTTTTGGAGTTATGACACCGGAGACCCCGTCCATCA 4380
Db      4321 GATGCACTTGAGCCCAAGCTTTTGGAGTTATGACACCGGAGACCCCGTCCATCA 4380
Qy      4381 TCTTGGCGGTCAATGCTTCTTATGAGTCTCTCTACATCGTGTGGAAGTGGCTCA 4440
Db      4381 TCTTGGCGGTCAATGCTTCTTATGAGTCTCTCTACATCGTGTGGAAGTGGCTCA 4440
Qy      4441 GGAGACACCGCAAGACCAAGAAAGTGAACAGACGAAAGCCGCTTGGCAAGACAG 4500
Db      4441 GGAGACACCGCAAGACCAAGAAAGTGAACAGACGAAAGCCGCTTGGCAAGACAG 4500
Qy      4501 CAATCAAGCTGCTGTAATGTCTGACAGAAAGCGGAGAAAGACAGGCACTTGTAACT 4560
Db      4501 CAATCAAGCTGCTGTAATGTCTGACAGAAAGCGGAGAAAGACAGGCACTTGTAACT 4560
Qy      4561 GCGCCGAGGACCAAGCGCGGACTTTGCGCTTTCGGAAGAGTCACTGGGCAAGCG 4620
Db      4561 GCGCCGAGGACCAAGCGCGGACTTTGCGCTTTCGGAAGAGTCACTGGGCAAGCG 4620
Qy      4621 CCGAGTGTGCGCGGTGGAGATCACTTGGGCAAGCCCATCTGCGCAAGCAAGTT 4680
Db      4621 CCGAGTGTGCGCGGTGGAGATCACTTGGGCAAGCCCATCTGCGCAAGCAAGTT 4680
Qy      4681 GCGCGAGGACCAAGCGCGGACTTTGCGCTTTCGGAAGAGTCACTGGGCAAGCG 4740
Db      4681 GCGCGAGGACCAAGCGCGGACTTTGCGCTTTCGGAAGAGTCACTGGGCAAGCG 4740
Qy      4741 TCGAAGCGGTGACCGGCGCGGCGGCTTTCGGAAGAGTCACTGGGCAAGCG 4800
Db      4741 TCGAAGCGGTGACCGGCGCGGCGGCTTTCGGAAGAGTCACTGGGCAAGCG 4800
Qy      4801 TCGAAGCGGTGACCGGCGCGGCGGCTTTCGGAAGAGTCACTGGGCAAGCG 4860
Db      4801 TCGAAGCGGTGACCGGCGCGGCGGCTTTCGGAAGAGTCACTGGGCAAGCG 4860
Qy      4861 GTCGAGGCTGTGATCAGCGCGGCTTTCGGAAGAGTCACTGGGCAAGCG 4920
Db      4861 GTCGAGGCTGTGATCAGCGCGGCTTTCGGAAGAGTCACTGGGCAAGCG 4920
Qy      4921 AAAGGCTACTCAAGAACTAAGAGGTGCTCAATCCACTCTTCTTCCGCGCGAGGCG 4980
Db      4921 AAAGGCTACTCAAGAACTAAGAGGTGCTCAATCCACTCTTCTTCCGCGCGAGGCG 4980
Qy      4981 GATTACAGGCGCGGACGAGTCCGAGTCCGCGCGCGGCTGCAAGATGAGCG 5040
Db      4981 GATTACAGGCGCGGACGAGTCCGAGTCCGCGCGCGGCTGCAAGATGAGCG 5040
Qy      5041 CGCTCATGTATTCAGAGCGGCTTTCGCTTTCGCGCGGCTGCAAGATGAGCG 5100
Db      5041 CGCTCATGTATTCAGAGCGGCTTTCGCTTTCGCGCGGCTGCAAGATGAGCG 5100
Qy      5101 CGCTCATGTATTCAGAGCGGCTTTCGCTTTCGCGCGGCTGCAAGATGAGCG 5160

```

```

Db      5101 CGCTCATGTATTCAGAGCGGCTTTCGCTTTCGCGCGGCTGCAAGATGAGCG 5160
Qy      5161 ATTAGTTTCTTGTTCGTAACCAAGTGGTCCCTCCATTGAGTAAAGACTTGGTAG 5220
Db      5161 ATTAGTTTCTTGTTCGTAACCAAGTGGTCCCTCCATTGAGTAAAGACTTGGTAG 5220
Qy      5221 TCTTCAAGTACTGTTGAGTCTGCTGGTTCGATTCATTCGCAAGCAAGAGGT 5280
Db      5221 TCTTCAAGTACTGTTGAGTCTGCTGGTTCGATTCATTCGCAAGCAAGAGGT 5280
Qy      5281 GCGCAACTAGTAGGCGCGCCCTGGGATACCA 5312
Db      5281 GCGCAACTAGTAGGCGCGCCCTGGGATACCA 5312

RESULT 2
US-09-991-262-41
; Sequence 41, Application US/09991262
; Publication No. US20030041349A1
; GENERAL INFORMATION:
; APPLICANT: Christlan, P. D., Gordon, K. H.J., Hanzlik, T. N.
; TITLE OF INVENTION: Insect Viruses and Their Uses in
; Protecting Plants
; NUMBER OF SEQUENCES: 57
; CORRESPONDENCE ADDRESS:
; ADDRESSES: Flehr Hohnbach Test Albritton & Herbert LLP
; STREET: Four Embarcadero Center, Suite 3400
; CITY: San Francisco
; STATE: California
; COUNTRY: United States
; ZIP: 94111-4187
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/991,262
; FILING DATE: 20-Nov. US20030041349A1-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/234,238
; FILING DATE: 20-Jan-1999
; APPLICATION NUMBER: US 08/485,355
; FILING DATE: 07-Jun-1995
; APPLICATION NUMBER: US 08/440,522
; FILING DATE: 12-May-1995
; APPLICATION NUMBER: US 08/089,372
; FILING DATE: 08-Jul-1993
; APPLICATION NUMBER: AU PL4081/92
; FILING DATE: 14-Aug-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Treacartin, Richard F.
; REGISTRATION NUMBER: 31,801
; REFERENCE/DOCKET NUMBER: A-58631-3/RFT/DSS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 781-1989
; TELEFAX: (415) 398-3249
; TELEX: 910 277299
; INFORMATION FOR SEQ ID NO: 41:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5312 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: unknown
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 4218..4514
; SEQUENCE DESCRIPTION: SEQ ID NO: 41:
US-09-991-262-41
Query Match 100.0%; Score 5312; DB 11; Length 5312;

```


QY 2161 CAATGGGTTTATGACTTCCAGGCACAGCGGAAACATGCCGCTCGTTCCGACGTC 2220
 DB 2161 CAATCGGGTTTATGACTTCCAGGCACAGCGGAAACATGCCGCTCGTTCCGACGTC 2220
 QY 2221 GTTAAAGCAGGCGCGTGGGCACTTTCAACCAACAGAGCGTGTCCGGCCGACGTCGTT 2280
 DB 2221 GTTAAAGCAGGCGCGTGGGCACTTTCAACCAACAGAGCGTGTCCGGCCGACGTCGTT 2280
 QY 2281 GCCACACGTTTTTCCAGAGCTTGTACCCCGGTGACAAACAACCTCAGGGTCCGTCGA 2340
 DB 2281 GCCACACGTTTTTCCAGAGCTTGTACCCCGGTGACAAACAACCTCAGGGTCCGTCGA 2340
 QY 2341 TCCATCAGCCACGCTCCCGCAGACTACCGCAACAGCGAGCGCAACCGCTTCCGTCAG 2400
 DB 2341 TCCATCAGCCACGCTCCCGCAGACTACCGCAACAGCGAGCGCAACCGCTTCCGTCAG 2400
 QY 2401 CAGGAGGAAAGTCGGGCAACCGGGGCTGAGGGGCGATGCTGTGACGAGCGGAGGA 2460
 DB 2401 CAGGAGGAAAGTCGGGCAACCGGGGCTGAGGGGCGATGCTGTGACGAGCGGAGGA 2460
 QY 2461 CGCACTTTTGGCTGTGCTATTCGATTAACAACGCTCCACAGCAGAGAGAAAGCTCTC 2520
 DB 2461 CGCACTTTTGGCTGTGCTATTCGATTAACAACGCTCCACAGCAGAGAGAAAGCTCTC 2520
 QY 2521 GCTGAGAGTCCGACCTTCTAGTCCGATCAACGCGCAACCAACCACTGTACATCCG 2580
 DB 2521 GCTGAGAGTCCGACCTTCTAGTCCGATCAACGCGCAACCAACCACTGTACATCCG 2580
 QY 2581 GACCCGACAGGTGACATTTAGAGAGCAACTCAACATAGCGGAAAGCGAGGTTTACA 2640
 DB 2581 GACCCGACAGGTGACATTTAGAGAGCAACTCAACATAGCGGAAAGCGAGGTTTACA 2640
 QY 2641 GACATCCCTGCAACCCCTGAGATCAAGACTGTCAAAACCGAGTGAAGGTGACGCAAC 2700
 DB 2641 GACATCCCTGCAACCCCTGAGATCAAGACTGTCAAAACCGAGTGAAGGTGACGCAAC 2700
 QY 2701 GAAATGATGCAACGATACCCCGCAGAGTCCACGCGCAACGAGCAATCACTGTCTC 2760
 DB 2701 GAAATGATGCAACGATACCCCGCAGAGTCCACGCGCAACGAGCAATCACTGTCTC 2760
 QY 2761 CGCAAGAACCTTGGGGACCAACCGGACTGTGCTGTGCTTGGCGAAGACCGGCTAC 2820
 DB 2761 CGCAAGAACCTTGGGGACCAACCGGACTGTGCTGTGCTTGGCGAAGACCGGCTAC 2820
 QY 2821 GAGGTGTTTGGCGGTGTGCTGCAAAATCAACGATGAGCTTGCAGAACCCGACGACCCG 2880
 DB 2821 GAGGTGTTTGGCGGTGTGCTGCAAAATCAACGATGAGCTTGCAGAACCCGACGACCCG 2880
 QY 2881 AAGCCGATAGGCGGTTCCAGAAAGGGGTACAGTGGGTCAACGAGCTGTAAAC 2940
 DB 2881 AAGCCGATAGGCGGTTCCAGAAAGGGGTACAGTGGGTCAACGAGCTGTAAAC 2940
 QY 2941 AAACACAGGCGGCTCCAGAGCTGTGCTCCGCTTACCAACGAGGAGCGGTGACCTGCCG 3000
 DB 2941 AAACACAGGCGGCTCCAGAGCTGTGCTCCGCTTACCAACGAGGAGCGGTGACCTGCCG 3000
 QY 3001 CTACAGAAAGCTTAAGAGAGAGCTCAACGAGCTTAACTCGCTTACCGACATTTGGAG 3060
 DB 3001 CTACAGAAAGCTTAAGAGAGAGCTCAACGAGCTTAACTCGCTTACCGACATTTGGAG 3060
 QY 3061 TGGACGTGTCACTGAGAGACGCGCGGTGACCGAGCTGTCTTGAAGACCAAGCTTCAACC 3120
 DB 3061 TGGACGTGTCACTGAGAGACGCGCGGTGACCGAGCTGTCTTGAAGACCAAGCTTCAACC 3120
 QY 3121 CAACGCGGCGGACCGCTGCAAGACTGTGCTGAGAGCAAGCAACCTTCAATCGGTGACATA 3180
 DB 3121 CAACGCGGCGGACCGCTGCAAGACTGTGCTGAGAGCAAGCAACCTTCAATCGGTGACATA 3180
 QY 3181 GACTTCTTATGAGAGACTGAGAGAAAGTGTCCGCCAAGCGATCAATACGGGCAAGGTC 3240
 DB 3181 GACTTCTTATGAGAGACTGAGAGAAAGTGTCCGCCAAGCGATCAATACGGGCAAGGTC 3240

QY 3241 GGGCAGGGGATGCGCGCTCACTCAAGTCTCAACTTGTCTCGCGCTTGGATACGC 3300
 DB 3241 GGGCAGGGGATGCGCGCTCACTCAAGTCTCAACTTGTCTCGCGCTTGGATACGC 3300
 QY 3301 ATACTGAGAGATATCTCCGTACCGGAGCCGACCGTCCGGTACAGCAACGGTCTCCC 3360
 DB 3301 ATACTGAGAGATATCTCCGTACCGGAGCCGACCGTCCGGTACAGCAACGGTCTCCC 3360
 QY 3361 GACGAAAGAGGCGCATGCTGTGAGAGGAGATCAATCAAGTCCACACGCGCATGTC 3420
 DB 3361 GACGAAAGAGGCGCATGCTGTGAGAGGAGATCAATCAAGTCCACACGCGCATGTC 3420
 QY 3421 GTCTGGGCGGACTGTGACCGAGTTTGAACCGCCCAATTAACGAGTGAAGCTCTTC 3480
 DB 3421 GTCTGGGCGGACTGTGACCGAGTTTGAACCGCCCAATTAACGAGTGAAGCTCTTC 3480
 QY 3481 GCCGCCCTTTTGAAGGCGATCGGACGCGCTGACGCTGAGTATATTAATTGAGAGACG 3540
 DB 3481 GCCGCCCTTTTGAAGGCGATCGGACGCGCTGACGCTGAGTATATTAATTGAGAGACG 3540
 QY 3541 TGTGGGAAACGACCTTTCGAGCGAAGGGTCTAGGCTCCGTTGAAGTCAAGGTCGTC 3600
 DB 3541 TGTGGGAAACGACCTTTCGAGCGAAGGGTCTAGGCTCCGTTGAAGTCAAGGTCGTC 3600
 QY 3601 GACTCCGGGCGAGCTTGAACCGCTTTCGCGCAACCACTTCTTCCGCGCATGCTC 3660
 DB 3601 GACTCCGGGCGAGCTTGAACCGCTTTCGCGCAACCACTTCTTCCGCGCATGCTC 3660
 QY 3661 ACGCTCTTCCGGGCGCTCAAGGTTCCGACTTCAAAAGGCGAAGCTCGCTCTGAGT 3720
 DB 3661 ACGCTCTTCCGGGCGCTCAAGGTTCCGACTTCAAAAGGCGAAGCTCGCTCTGAGT 3720
 QY 3721 AGCATTACCTTCGTTTGAAGGCTAGCGCTTCAACATGAGGCGAAGCTTACAAAGCAAA 3780
 DB 3721 AGCATTACCTTCGTTTGAAGGCTAGCGCTTCAACATGAGGCGAAGCTTACAAAGCAAA 3780
 QY 3781 CATTTGAAGTGAAGTGAAGAAATCGTCCGTACATGGACTCCTCGTCCGCTGAG 3840
 DB 3781 CATTTGAAGTGAAGTGAAGAAATCGTCCGTACATGGACTCCTCGTCCGCTGAG 3840
 QY 3841 CAGTGTCTCGAACCTGTCAAGAGCGCTCAAGATATTTGGGCGCTTACACAAGC 3900
 DB 3841 CAGTGTCTCGAACCTGTCAAGAGCGCTCAAGATATTTGGGCGCTTACACAAGC 3900
 QY 3901 GAACTCTTTTACTCCAGATGCTGAGAGGCTGTGAGAGCATCAACAGGCTGAGTGAAC 3960
 DB 3901 GAACTCTTTTACTCCAGATGCTGAGAGGCTGTGAGAGCATCAACAGGCTGAGTGAAC 3960
 QY 3961 GCCCGTACCAAGCTCTCTGTGCAATGTGAGAGCATCAACAGGCTGAGTGAAC 4020
 DB 3961 GCCCGTACCAAGCTCTCTGTGCAATGTGAGAGCATCAACAGGCTGAGTGAAC 4020
 QY 4021 TCTGCGGCGTACATCAAGAGCTGTGTTGCTTGGGCGGCGAGCATTTCCGTTTGA 4080
 DB 4021 TCTGCGGCGTACATCAAGAGCTGTGTTGCTTGGGCGGCGAGCATTTCCGTTTGA 4080
 QY 4081 CAATGCGGCGTGTGCTGAGAGGCTGAGAGCAACCGGACGCTTACAGAGCATATCCG 4140
 DB 4081 CAATGCGGCGTGTGCTGAGAGGCTGAGAGCAACCGGACGCTTACAGAGCATATCCG 4140
 QY 4141 GCTAACGTGCGGAGATGCTGTGCTTGAACAGCTTGTGAAGCCCGGACAGGCGCGCCG 4200
 DB 4141 GCTAACGTGCGGAGATGCTGTGCTTGAACAGCTTGTGAAGCCCGGACAGGCGCGCCG 4200
 QY 4201 GCAAGTTTCTTTCGACATGTGAGAGGCGGAAACGCTTCTTCACTTACCGGAAAGCT 4260
 DB 4201 GCAAGTTTCTTTCGACATGTGAGAGGCGGAAACGCTTCTTCACTTACCGGAAAGCT 4260
 QY 4261 GGTGTTTCTGCACTAACAGCAAGCTTTCGAGCTGCGGACGCTTCATGG 4320
 DB 4261 GGTGTTTCTGCACTAACAGCAAGCTTTCGAGCTGCGGACGCTTCATGG 4320
 QY 4321 GATGCACTGAGCGCAACAGCTTTTCCAGTTATTAACACCGGAGACCCCGTCCATCA 4380

Db	4321	GATGACCTTGACGCAACAGCTTTTGGAGTTATTGACACCGAGAGACCCCGTCCATCATCA	4380
Qy	4381	TCCTCGCCGTCATTCGTCTTTCATCGGACTCCTCTCATTCGTGTGGAAAGTGGCTCAAGTGT	4440
Db	4381	TCCTCGCCGTCATTCGTCTTTCATCGGACTCCTCTCATTCGTGTGGAAAGTGGCTCAAGTGT	4440
Qy	4441	GGAGACACCGCAAGAGACCAAGAGACTTGGAAACGACAGAAAAGCGCCTTCCGAAGACAG	4500
Db	4441	GGAGACACCGCAAGAGACCAAGAGACTTGGAAACGACAGAAAAGCGCCTTCCGAAGACAG	4500
Qy	4501	CAATCAGCTCGTCTGTGAATGTCTGGACAGAAAGCGAGAAAAGACAGCAAGTTCGTTAACT	4560
Db	4501	CAATCAGCTCGTCTGTGAATGTCTGGACAGAAAGCGAGAAAAGACAGCAAGTTCGTTAACT	4560
Qy	4561	GGCCCCACTGCTCCGAGAGCCCTCAATTCATTTTGGAAAAGAGTCGACTGGAGACCGAG	4620
Db	4561	GGCCCCACTGCTCCGAGAGCCCTCAATTCATTTTGGAAAAGAGTCGACTGGAGACCGAG	4620
Qy	4621	CCGACTGTGCGCGCTGGGACATCACTTCGAGCAACCCCATCTCTGCGCACAGGACAGATT	4680
Db	4621	CCGACTGTGCGCGCTGGGACATCACTTCGAGCAACCCCATCTCTGCGCACAGGACAGATT	4680
Qy	4681	GCCCGCAGAGACCAAGCCGGAGATTGGCCGCTTTCCTGGGTTCCCAAGTCTGCGCCGTCGTCTC	4740
Db	4681	GCCCGCAGAGACCAAGCCGGAGATTGGCCGCTTTCCTGGGTTCCCAAGTCTGCGCCGTCGTCTC	4740
Qy	4741	TGCAAGCGCTGACCGGCCCCCACCACACTGCCGTTGGAAAAGAGTCAACCCGCTCCACGCG	4800
Db	4741	TGCAAGCGCTGACCGGCCCCCACCACACTGCCGTTGGAAAAGAGTCAACCCGCTCCACGCG	4800
Qy	4801	TGGAAGGCGCTGACCGGAGACCGAACGCGAGATCAAGGAGAGACCGGAGACAGCGGCGGTC	4860
Db	4801	TGGAAGGCGCTGACCGGAGACCGAACGCGAGATCAAGGAGAGACCGGAGACAGCGGCGGTC	4860
Qy	4861	GTCACAGGCTCGATCAGCGGCGCGTTATCTCAGAAAGACGAACTTTCCTCCGACGCAATCC	4920
Db	4861	GTCACAGGCTCGATCAGCGGCGCGTTATCTCAGAAAGACGAACTTTCCTCCGACGCAATCC	4920
Qy	4921	AAAGGCTACTCAAGAACTTAAGGAGTGTCTCACAATTCACCTCTTTTCTGCCCCAGATCGC	4980
Db	4921	AAAGGCTACTCAAGAACTTAAGGAGTGTCTCACAATTCACCTCTTTTCTGCCCCAGATCGC	4980
Qy	4981	GATTACCAAGGCGCGGCACTGCGCAACAGTCCGAGTCTGCGCGCGCGCTCGAGAGATGGCG	5040
Db	4981	GATTACCAAGGCGCGGCACTGCGCAACAGTCCGAGTCTGCGCGCGCGCTCGAGAGATGGCG	5040
Qy	5041	CGCTCATGTATTACACAGAGCGTTGAGCTTCATCTGCGCGCACTGCGCACTTGAACGCGATA	5100
Db	5041	CGCTCATGTATTACACAGAGCGTTGAGCTTCATCTGCGCGCACTGCGCACTTGAAGGCGATA	5100
Qy	5101	CGCTCTACCTTGGAACTCTGTTCCCGATGTAAAGATCAGCAAGAGCGCATGAAAGAACAA	5160
Db	5101	CGCTCTACCTTGGAACTCTGTTCCCGATGTAAAGATCAGCAAGAGCGCATGAAAGAACAA	5160
Qy	5161	ATTAGTTTCCCTGTTGCTGTAACAAGAGTGTCCCTCCCATTAAGGTAAGAACTCTGTGTAG	5220
Db	5161	ATTAGTTTCCCTGTTGCTGTAACAAGAGTGTCCCTCCCATTAAGGTAAGAACTCTGTGTAG	5220
Qy	5221	TCCTCAACGTTACTCGTTGAGTCTGTGCTGCGGTTGCAATTCATTCCTCAAGCAGCAAAAGGTT	5280
Db	5221	TCCTCAACGTTACTCGTTGAGTCTGTGCTGCGGTTGCAATTCATTCCTCAAGCAGCAAAAGGTT	5280
Qy	5281	GCGCAACTAGTACGCGCGCGCCCTGGGAGTAACA 5312	
Db	5281	GCGCAACTAGTACGCGCGCGCCCTGGGAGTAACA 5312	

```

RESULT 3
US-09-991-262-43
; Sequence 43, Application US/09991262
; Publication No. US20030041349A1
; GENERAL INFORMATION:

```

```

1  APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
2  TITLE OF INVENTION: Insect Viruses and Their Uses in
3  Protecting Plants
4  NUMBER OF SEQUENCES: 57
5  CORRESPONDENCE ADDRESSES:
6  ADDRESSEE: Flehr Hohbach Test Albritton & Herbert LLP
7  STREET: Four Embarcadero Center, Suite 3400
8  CITY: San Francisco
9  STATE: California
10 COUNTRY: United States
11 ZIP: 94111-4187
12 COMPUTER READABLE FORM:
13 MEDIUM TYPE: Floppy disk
14 COMPUTER: IBM PC compatible
15 OPERATING SYSTEM: PC-DOS/MS-DOS
16 SOFTWARE: PatentIn Release #1.0, Version #1.30
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/09/991,262
19 FILING DATE: 20-No. US20030041349A1-2001
20 CLASSIFICATION: <Unknown>
21 PRIOR APPLICATION DATA:
22 APPLICATION NUMBER: US/09/234,238
23 FILING DATE: 20-JAN-1999
24 APPLICATION NUMBER: US 08/485,355
25 FILING DATE: 07-JUN-1995
26 APPLICATION NUMBER: US 08/440,522
27 FILING DATE: 12-MAY-1995
28 APPLICATION NUMBER: US 08/089,372
29 FILING DATE: 08-JUL-1993
30 APPLICATION NUMBER: AU PL4081/92
31 FILING DATE: 14-AUG-1992
32 ATTORNEY/AGENT INFORMATION:
33 NAME: Treacartin, Richard F.
34 REGISTRATION NUMBER: 31,801
35 REFERENCE/DOCKET NUMBER: A-58631-3/REF/DSS
36 TELECOMMUNICATION INFORMATION:
37 TELEPHONE: (415) 781-1989
38 TELEFAX: (415) 398-3249
39 TELEX: 910 277299
40 INFORMATION FOR SEQ ID NO: 43:
41 SEQUENCE CHARACTERISTICS:
42 LENGTH: 5312 base pairs
43 TYPE: nucleic acid
44 STRANDEDNESS: unknown
45 TOPOLOGY: unknown
46 MOLECULE TYPE: DNA
47 FEATURE:
48 NAME/KEY: CDS
49 LOCATION: 4518..4937
50 US-09-991-262-43 SEQUENCE DESCRIPTION: SEQ ID NO: 43:
51
52 Query Match 100.0%; Score 5312; DB 11; Length 5312;
53 Best Local Similarity 100.0%; Pred. No. 0;
54 Matches 5312; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
55
56 QY 1 GTTCTGCTCTCCCGGACGGTAAATATATAGGGGAACTATGTACGGGAAAGCGACAGACGTG 60
57 Db 1 GTTCTGCTCTCCCGGACGGTAAATATATAGGGGAACTATGTACGGGAAAGCGACAGACGTG 60
58
59 QY 61 GCGGCTGTCTACGCCGGGCGAGATGTGGCTTACCGGCAAGTATCTGCAGACAGAGCAATC 120
60 Db 61 GCGGCTGTCTACGCCGGGCGAGATGTGGCTTACCGGCAAGTATCTGCAGACAGAGCAATC 120
61 GCGGCTGTCTACGCCGGGCGAGATGTGGCTTACCGGCAAGTATCTGCAGACAGAGCAATC 120
62
63 QY 121 AAGTTGACCTTGGCCCGGCGACCTAAGGACCTAGAAACCTTCCACAGACTGTACTATCCG 180
64 Db 121 AAGTTGACCTTGGCCCGGCGACCTAAGGACCTAGAAACCTTCCACAGACTGTACTATCCG 180
65
66 QY 181 CTGGCGTTCAAAAGGGGCACTTTTACCCCGGACACAACAACCGGATCTGGCCGGGACCAA 240
67 Db 181 CTGGCGTTCAAAAGGGGCACTTTTACCCCGGACACAACAACCGGATCTGGCCGGGACCAA 240
68
69 QY 241 CGTGTCCGAAGAGGTTTCTGCACAAATTTCCGCCAGGGGACGTAGACAGTGTGAGATA 300

```

Db	241	CGTGTGCAAGAGAGTTCTTGCAACAATTTGGCAGGGGAGCGTAGACACAGTGTCTGAGATA	300
Qy	301	GGGCGCGTCTCTGCAACAGCGCACTTTAAGCTCAATGAGGAGCAACGAAACGCCCGCTCGACAGAC	360
Db	301	GGGCGCGTCTCTGCAACAGCGCACTTTAAGCTCAATGAGGAGCAACGAAACGCCCGCTCGACAGAC	360
Qy	361	TATCAACGGGTGCAACAAAGTACGGGCAACCCCGGACGGCTCGCGACACATTAACGCGCTTAAGAG	420
Db	361	TATCAACGGGTGCAACAAAGTACGGGCAACCCCGGACGGCTCGCGACACATTAACGCGCTTAAGAG	420
Qy	421	TCTAGATCCGTCGCGCAACAGCGCGGCGCGGAGTTCAAGGCGCAACGCGCTCACTGCTCGCCAAC	480
Db	421	TCTAGATCCGTCGCGCAACAGCGCGGCGCGGAGTTCAAGGCGCAACGCGCTCACTGCTCGCCAAC	480
Qy	481	GGCATTTGCTCCCGCACCTTCTGCGTGCAGCGAGATCGAGCTTCTTTCGCGGTTCAAAATCGCGC	540
Db	481	GGCATTTGCTCCCGCACCTTCTGCGTGCAGCGAGATCGAGCTTCTTTCGCGGTTCAAAATCGCGC	540
Qy	541	GTTTGAATTGCGCAATCACTCCCTCTATGACGTGACCCCTAGAGAGACTGAGCAATGCGTTT	600
Db	541	GTTTGAATTGCGCAATCACTCCCTCTATGACGTGACCCCTAGAGAGACTGAGCAATGCGTTT	600
Qy	601	GAGAACCAACGAGACTTTCACATGCTCCGCGCGTTTACATGCACATGCCAAGAAAGCTGCTTAC	660
Db	601	GAGAACCAACGAGACTTTCACATGCTCCGCGCGTTTACATGCACATGCCAAGAAAGCTGCTTAC	660
Qy	661	ATGGACAAACGGGTAAATGACCGAGCTGGGCTACCGGCTTCCAGCGTTATTTGAAGAGCTATG	720
Db	661	ATGGACAAACGGGTAAATGACCGAGCTGGGCTACCGGCTTCCAGCGTTATTTGAAGAGCTATG	720
Qy	721	GCTGTGAAGACCTGCGCAATTCACAGGGGGAGGACCTCCGCTCCCACTTCCCTGAGTTGAGAC	780
Db	721	GCTGTGAAGACCTGCGCAATTCACAGGGGGAGGACCTCCGCTCCCACTTCCCTGAGTTGAGAC	780
Qy	781	TTTCATCAACGAGAGCCCAAGAGCGGCGCATGAGAGGCTGAGCGCGCGGCTCCTACTCC	840
Db	781	TTTCATCAACGAGAGCCCAAGAGCGGCGCATGAGAGGCTGAGCGCGCGGCTCCTACTCC	840
Qy	841	AGAGCGCGCGCATTTTCTCCGGGAGACGACGACGTGGGGTGAATGGGTACTTACAGACATTTC	900
Db	841	AGAGCGCGCGCATTTTCTCCGGGAGACGACGACGTGGGGTGAATGGGTACTTACAGACATTTC	900
Qy	901	CACACATGAGTCTGCGCTCACTAAGTGTGAGAACTAACCCCACTCCGTTTGTTTTCTCACTC	960
Db	901	CACACATGAGTCTGCGCTCACTAAGTGTGAGAACTAACCCCACTCCGTTTGTTTTCTCACTC	960
Qy	961	CATATAGAGTCCAGAGGCGCCACGAGCTCCAGCATTTGAGCTGCGCATCACTCGCGCGCA	1020
Db	961	CATATAGAGTCCAGAGGCGCCACGAGCTCCAGCATTTGAGCTGCGCATCACTCGCGCGCA	1020
Qy	1021	CGTGAAGACCGGCAATGCTGGCGGTGCTGCCAAGAGACCTCCAAAGGCGCTGTGAGAAATCCCA	1080
Db	1021	CGTGAAGACCGGCAATGCTGGCGGTGCTGCCAAGAGACCTCCAAAGGCGCTGTGAGAAATCCCA	1080
Qy	1081	AACATCTTTTATATAGCGCGAGCGGTGCGGCACTAGACATAGACCATCTTACGTCAAG	1140
Db	1081	AACATCTTTTATATAGCGCGAGCGGTGCGGCACTAGACATAGACCATCTTACGTCAAG	1140
Qy	1141	CACAAAGTCAACATGCTGCTCAATTTATATGCAAAACGCGTCTGAGAAAGAACTAGTGCAC	1200
Db	1141	CACAAAGTCAACATGCTGCTCAATTTATATGCAAAACGCGTCTGAGAAAGAACTAGTGCAC	1200
Qy	1201	ATGACCCGCTTGATGTGCTTGGCGCGCGCTTAGCTGCGCGGAGATCTGTGTGCTTACGAA	1260
Db	1201	ATGACCCGCTTGATGTGCTTGGCGCGCGCTTAGCTGCGCGGAGATCTGTGTGCTTACGAA	1260
Qy	1261	GTCACCGAGAGCTCCTGGAGACATCTCAACCGGCTGACCTGTCCGACCTGTGTGTCTT	1320
Db	1261	GTCACCGAGAGCTCCTGGAGACATCTCAACCGGCTGACCTGTGTCCGACCTGTGTGTCTT	1320
Qy	1321	TACGTCCTCCACATCATGAGAGCGGAAGGCTGCGGTGCTGTACAGACCGCCAGAGAC	1380

Db	1321	TAGGTCTCCACATCATTCAGAGGCCGGAAGGGCTGCGGTGCGCTGTCAAGACCGCCAAAGAC	1380
Qy	1381	GACGCTCTTTGAGAGAGACTTCGTGTTCTGAGAGAGTCTCAAGCAAGCTCTTGAGCTCTCTGTTGC	1440
Db	1381	GACGCTCTTTGAGAGAGACTTCGTGTTCTGAGAGAGTCTCAAGCAAGCTCTTGAGCTCTCTGTTGC	1440
Qy	1441	GCTCTGGCGAACCTTCGAAGGCAACGATGCTGTTTATCTTAAGAGGCGCTGCTGATTAATAC	1500
Db	1441	GCTCTGGCGAACCTTCGAAGGCAACGATGCTGTTTATCTTAAGAGGCGCTGCTGATTAATAC	1500
Qy	1501	CGAGTCCACTGCTCGAGAGACATTAATCTGAGAGCTGCGCGCTGTCCTCCCTGAACAGGTGGC	1560
Db	1501	CGAGTCCACTGCTCGAGAGACATTAATCTGAGAGCTGCGCGCTGTCCTCCCTGAACAGGTGGC	1560
Qy	1561	TTCTGCGCGTCCCGCGTACCAAGCTGCGCGCTCTTTCAGCAAGAGGAAGCTTGAAGTTC	1620
Db	1561	TTCTGCGCGTCCCGCGTACCAAGCTGCGCGCTCTTTCAGCAAGAGGAAGCTTGAAGTTC	1620
Qy	1621	CTTGCGCGAAGCTGGCTGCTACAGAAAGCTGCGGTACTTCCACTCTCTCTGTGAGAGAG	1680
Db	1621	CTTGCGCGAAGCTGGCTGCTACAGAAAGCTGCGGTACTTCCACTCTCTCTGTGAGAGAG	1680
Qy	1681	CCCCAAGGTTTCAGAGCGCGAATTGTGACAGCGAGCGGCGTCACTCCCGAATACCGC	1740
Db	1681	CCCCAAGGTTTCAGAGCGCGAATTGTGACAGCGAGCGGCGTCACTCCCGAATACCGC	1740
Qy	1741	GCCAATTGCAAGAGGTCTTCACACCGACGTCAGAGAGCTCAAGATCAACCTTCGAGAAC	1800
Db	1741	GCCAATTGCAAGAGGTCTTCACACCGACGTCAGAGAGCTCAAGATCAACCTTCGAGAAC	1800
Qy	1801	GCCTCTAAGACATTCAGACGGGCTCAACCTCTGCCCACTCAGAGGCTTCGAGATGTACGAG	1860
Db	1801	GCCTCTAAGACATTCAGACGGGCTCAACCTCTGCCCACTCAGAGGCTTCGAGATGTACGAG	1860
Qy	1861	GGCCCGCAGGCAAGCGGCAAGAGCGGAGCACTTCATGCGCGCCCTTGAAGGCGCGGGCGGT	1920
Db	1861	GGCCCGCAGGCAAGCGGCAAGAGCGGAGCACTTCATGCGCGCCCTTGAAGGCGCGGGCGGT	1920
Qy	1921	AAAGCACTTTATCGTGGCACCCACACAGAGAACTGAGAGAGGCTATGAGACCGCGGATCAAA	1980
Db	1921	AAAGCACTTTATCGTGGCACCCACACAGAGAACTGAGAGAGGCTATGAGACCGCGGATCAAA	1980
Qy	1981	CCGCCGTGCGCTCGGCTACCGCAACATGTCGCCCTTGCGAATTCCTCGCGTGCACCGCC	2040
Db	1981	CCGCCGTGCGCTCGGCTACCGCAACATGTCGCCCTTGCGAATTCCTCGCGTGCACCGCC	2040
Qy	2041	GAGGGCGGCGCTTTCGCTACCGTGGTTATGCAAGAGTCTTATGATTCCTCCGCTGTGAC	2100
Db	2041	GAGGGCGGCGCTTTCGCTACCGTGGTTATGCAAGAGTCTTATGATTCCTCCGCTGTGAC	2100
Qy	2101	GTCGGATCTGTCAGCGCTTGTCCCGGAGCTCAAGAAATGCTCTGTGAGGGAAGTCCAC	2160
Db	2101	GTCGGATCTGTCAGCGCTTGTCCCGGAGCTCAAGAAATGCTCTGTGAGGGAAGTCCAC	2160
Qy	2161	CAAAATCGGTTTATAGACTTTCGAAGGCAAGAGCGCAATGCGCTGTTTCGCGACGTC	2220
Db	2161	CAAAATCGGTTTATAGACTTTCGAAGGCAAGAGCGCAATGCGCTGTTTCGCGACGTC	2220
Qy	2221	GTTAAGCAGTGCCTGCGCGGCACTTTCGAACCAAAACCAAGCGCTGTCCGGCGGACGTCGT	2280
Db	2221	GTTAAGCAGTGCCTGCGCGGCACTTTCGAACCAAAACCAAGCGCTGTCCGGCGGACGTCGT	2280
Qy	2281	GCACACAGTTCCTTCAGAGCTTGTACCCCGGGTGCACACCACTCAAGGTTGCGTCCGCA	2340
Db	2281	GCACACAGTTCCTTCAGAGCTTGTACCCCGGGTGCACACCACTCAAGGTTGCGTCCGCA	2340
Qy	2341	TCGATCAGCCACGTGCGCCGACGTAACGCAACGCAAGCGCAAAAGCTCTGCTTACG	2400
Db	2341	TCGATCAGCCACGTGCGCCGACGTAACGCAACGCAAGCGCAAAAGCTCTGCTTACG	2400
Qy	2401	CAGAGAGAAAAGTGGCGGCAAGGGGCTGAAGGGCGCGATGATCTGTGACGAAGCCCAAGGA	2460
Db	2401	CAGAGAGAAAAGTGGCGGCAAGGGGCTGAAGGGCGCGATGATCTGTGACGAAGCCCAAGGA	2460


```

Db 4621 CGGACGTGCGCGTGCAGATCATTCTTGGCAACCCCATCTGCGCCAGCAAGGTT 4680
Qy 4681 GCGCGAGAGACACGCGCGGACTTTTGGCGCTTCTCTGGGTTCCAGTCTGCGCGTGTCTC 4740
Db 4681 GCGCGAGAGACACGCGCGGACTTTTGGCGCTTCTCTGGGTTCCAGTCTGCGCGTGTCTC 4740
Qy 4741 TCGAAGCCGTTACCGGCGCCCGCCAGACTGCCGTTTGAAGAAAGTCAACCCGCTCCACGG 4800
Db 4741 TCGAAGCCGTTACCGGCGCCCGCCAGACTGCCGTTTGAAGAAAGTCAACCCGCTCCACGG 4800
Qy 4801 TCGAAGGGGGTGAACCGGAGACCGGAGAGTGAAGGAGAGACCCGGAAGACAGGGGGGTC 4860
Db 4801 TCGAAGGGGGTGAACCGGAGACCGGAGAGTGAAGGAGAGACCCGGAAGACAGGGGGGTC 4860
Qy 4861 GTCCAGGCTCTGATCAGCGCGCGCTTATCTCAGAAAGAGAGAGTCTTCTCCAGCATCC 4920
Db 4861 GTCCAGGCTCTGATCAGCGCGCGCTTATCTCAGAAAGAGAGAGTCTTCTCCAGCATCC 4920
Qy 4921 AAAAGCTACTCAAGAACTAAAGGATGCTCAAAATCCACTTTTCTCTGCGCGAGTGG 4980
Db 4921 AAAAGCTACTCAAGAACTAAAGGATGCTCAAAATCCACTTTTCTCTGCGCGAGTGG 4980
Qy 4981 GATTACAGGCGCGCGGACCTGCGAGACAGTCCGAGTCTGCGCGCGCGCTGAGAGATGCG 5040
Db 4981 GATTACAGGCGCGCGGACCTGCGAGACAGTCCGAGTCTGCGCGCGCGCTGAGAGATGCG 5040
Qy 5041 CGCTCATGTATTCACAGACCGGCTTGTGCTTATCTGCGCGCGAGTCCGATGAAAGCGATA 5100
Db 5041 CGCTCATGTATTCACAGACCGGCTTGTGCTTATCTGCGCGCGAGTCCGATGAAAGCGATA 5100
Qy 5101 CGCTCATGTATTCACAGACCGGCTTGTGCTTATCTGCGCGCGAGTCCGATGAAAGCGATA 5160
Db 5101 CGCTCATGTATTCACAGACCGGCTTGTGCTTATCTGCGCGCGAGTCCGATGAAAGCGATA 5160
Qy 5161 ATTAGTTTCTTGTGTGTAACAAGTGTGCTTCCATTCATGAGTAAGCTCTGCTGAG 5220
Db 5161 ATTAGTTTCTTGTGTGTAACAAGTGTGCTTCCATTCATGAGTAAGCTCTGCTGAG 5220
Qy 5221 TCCTCAACGTTACGTTGAGTCTGCTGCGGTTGATTCATTCCCAAGAGAAAGGT 5280
Db 5221 TCCTCAACGTTACGTTGAGTCTGCTGCGGTTGATTCATTCCCAAGAGAAAGGT 5280
Qy 5281 GCGCAACTAGTACGCGCGCGCTTGGGATACCA 5312
Db 5281 GCGCAACTAGTACGCGCGCGCTTGGGATACCA 5312

```

```

RESULT 4
US-09-991-262-45
; Sequence 45, Application US/09991262
; Publication No. US20030041349A1
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H. J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Virtues and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESS: Flehr Hobach Test Albritton & Herbert LLP
STREET: Four Embardadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/991, 262
FILING DATE: 20-No. US20030041349A1-2001
CLASSIFICATION: <Unknown>

```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/234,238
FILING DATE: 20-JAN-1999
APPLICATION NUMBER: US 08/485,355
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacart, Richard P.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-3/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 5368 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 4944..5162
SEQUENCE DESCRIPTION: SEQ ID NO: 45:
US-09-991-262-45
Query Match 98.0%; Score 5206; DB 11; Length 5368;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 5312; Conservative 0; Mismatches 0; Indels 56; Gaps 5;

```

```

Qy 1 GTTTCGCTTCCCGGACCGTAAATATAGGGAACAAATTAAGCGCAACAGACGTC 60
Db 1 GTTTCGCTTCCCGGACCGTAAATATAGGGAACAAATTAAGCGCAACAGACGTC 60
Qy 61 GCGCGTGTCTACGCGCGGACAGATGTCCTTACGCGCAAGTACTGAGAGAGAGTC 120
Db 61 GCGCGTGTCTACGCGCGGACAGATGTCCTTACGCGCAAGTACTGAGAGAGAGTC 120
Qy 121 AAGTTGACTTGGCCCGGCACTGAAGGCACTGAAGAACTCTTCAAGACTGTATCCG 180
Db 121 AAGTTGACTTGGCCCGGCACTGAAGGCACTGAAGAACTCTTCAAGACTGTATCCG 180
Qy 181 CTGCGCTTCAAGAGGGGCACTTACCCCGGACACCAACCCGATCTTGGCCGGGACCA 240
Db 181 CTGCGCTTCAAGAGGGGCACTTACCCCGGACACCAACCCGATCTTGGCCGGGACCA 240
Qy 241 CGTGTGCAAGAAAGGTTCTGCAATTTTCCGAGGGGAGTGAAGCAAGTCTGAGATA 300
Db 241 CGTGTGCAAGAAAGGTTCTGCAATTTTCCGAGGGGAGTGAAGCAAGTCTGAGATA 300
Qy 301 GGGCGGTCTCTGCAAGCGCACTTAAGCTAAGTGGGCAACGAAACCCCGGTGCGAGAC 360
Db 301 GGGCGGTCTCTGCAAGCGCACTTAAGCTAAGTGGGCAACGAAACCCCGGTGCGAGAC 360
Qy 361 TATCAGGGGTGCAACGAATACGCAACCGGAGACGGCTTGGGACAACTTAAGGCTTAGAG 420
Db 361 TATCAGGGGTGCAACGAATACGCAACCGGAGACGGCTTGGGACAACTTAAGGCTTAGAG 420
Qy 421 TTAGATCCGTGCGCAAGGCGCGCGGAGTTCAAGGCGACGCTCACTGCTGCGCAAC 480
Db 421 TTAGATCCGTGCGCAAGGCGCGCGGAGTTCAAGGCGACGCTCACTGCTGCGCAAC 480
Qy 481 GGAATGCTCTCCGCACTTTTGTGCTGAGAGAGTGGCTTTTGGCGGTTCAATGCGGC 540
Db 481 GGAATGCTCTCCGCACTTTTGTGCTGAGAGAGTGGCTTTTGGCGGTTCAATGCGGC 540
Qy 541 GTTGAATGCCAATACATCCCTTATGACGTGACCCCTAGAGAGAGCTGGCAATGGTTT 600

```

Db 541 GTTGAATTCGCAATCACTCCCTCTATGACGTGACCCCTTAGAGAGACTGGCCAAATGCGTT 600
 Qy 601 GAGAACCAAGCACTTCAATGATGTCGCGCTTTCATGACATGCTCAGAGAGAGCTGCTTAC 660
 Db 601 GAGAACCAAGCACTTCAATGATGTCGCGCTTTCATGACATGCTCAGAGAGAGCTGCTTAC 660
 Qy 661 ATGGAACAAGCTGTTAATGCGGAGCTCGGCTTACCGCTTCCAGTTAATGAAAGCTAAG 720
 Db 661 ATGGAACAAGCTGTTAATGCGGAGCTCGGCTTACCGCTTCCAGTTAATGAAAGCTAAG 720
 Qy 721 GCTGGAAGAGACTGCGCATTCAGAGGAGGAGACTCCGCTCCACTTCCTGAGTTGAG 780
 Db 721 GCTGGAAGAGACTGCGCATTCAGAGGAGGAGACTCCGCTCCACTTCCTGAGTTGAG 780
 Qy 781 TTTCATCAACGAGAGCCCAAGAGCGGCGCATGAGAGGCTTGGCCGCGGCTCTACTCC 840
 Db 781 TTTCATCAACGAGAGCCCAAGAGCGGCGCATGAGAGGCTTGGCCGCGGCTCTACTCC 840
 Qy 841 AGAGCGCGCTCATTTTCTCCGCGCAGACGACTGGGGTATGCTGATCTTACAGACTTC 900
 Db 841 AGAGCGCGCTCATTTTCTCCGCGCAGACGACTGGGGTATGCTGATCTTACAGACTTC 900
 Qy 901 CACACATGGCTGCGCTACCTACCTAGGAGAACTACCCCACTCCGTTTGTTTCTCACTC 960
 Db 901 CACACATGGCTGCGCTACCTACCTAGGAGAACTACCCCACTCCGTTTGTTTCTCACTC 960
 Qy 961 CATATAGAAGTCAGAGGCGCCAGCGCTTCAGACATTTAGACTGCTGACATCGCGGCCA 1020
 Db 961 CATATAGAAGTCAGAGGCGCCAGCGCTTCAGACATTTAGACTGCTGACATCGCGGCCA 1020
 Qy 1021 CCTGAGAACCGCATGCTGCGCTGCTCCAGAGAGCTGCCAAGGCTCTGCAAGAACTCCA 1080
 Db 1021 CCTGAGAACCGCATGCTGCGCTGCTCCAGAGAGCTGCCAAGGCTCTGCAAGAACTCCA 1080
 Qy 1081 AACATCTTTTATTTAAGCGCGAGCGTGGGACATGAGCATTAAGCACTCTTACGTACAG 1140
 Db 1081 AACATCTTTTATTTAAGCGCGAGCGTGGGACATGAGCATTAAGCACTCTTACGTACAG 1140
 Qy 1141 CACAAAGTCACATGCTGCTCAATTTTATGCAAAAGCGCTCTGAGAGAGAACTAGTCAG 1200
 Db 1141 CACAAAGTCACATGCTGCTCAATTTTATGCAAAAGCGCTCTGAGAGAGAACTAGTCAG 1200
 Qy 1201 ATGACCGCTTATGCTGCTGCGCGCGCTGAGGCTGCGCGCATGCTGCTCGCTCAGAA 1260
 Db 1201 ATGACCGCTTATGCTGCTGCGCGCGCTGAGGCTGCGCGCATGCTGCTCGCTCAGAA 1260
 Qy 1261 GTTACCGAGAGCTCCTGGAACATCTCAAGGCTGACCTGCTCGGCACTGCTGCTCTT 1320
 Db 1261 GTTACCGAGAGCTCCTGGAACATCTCAAGGCTGACCTGCTCGGCACTGCTGCTCTT 1320
 Qy 1321 TACGCTCTTCACATCATCAAGGCGCGGAGGCTGCGCTGCTCAAGAACCGCGCAAGAAC 1380
 Db 1321 TACGCTCTTCACATCATCAAGGCGCGGAGGCTGCGCTGCTCAAGAACCGCGCAAGAAC 1380
 Qy 1381 GACGCTCTTGAAGAGCTTCTGTTCTGGAAGCTTCAAGCAAGCTTGGGCTCTCTGTTGC 1440
 Db 1381 GACGCTCTTGAAGAGCTTCTGTTCTGGAAGCTTCAAGCAAGCTTGGGCTCTCTGTTGC 1440
 Qy 1441 GGTTCGCGCAACTCTGAAAGGACCGGAGCTGCTTAACTAAAGGCGCTCCGTAAGAGTAC 1500
 Db 1441 GGTTCGCGCAACTCTGAAAGGACCGGAGCTGCTTAACTAAAGGCGCTCCGTAAGAGTAC 1500
 Qy 1501 CGAGTCACTCGCTCGGAGACATAATCTGAGAGCTGCGCTGCTCCCTGAACAGAGTCCGC 1560
 Db 1501 CGAGTCACTCGCTCGGAGACATAATCTGAGAGCTGCGCTGCTCCCTGAACAGAGTCCGC 1560
 Qy 1561 TTTCCTGCGCTCCGCGCTACCACTGCGCGCTTTCACGACAGGAGAAAGCTTGAAGTTC 1620
 Db 1561 TTTCCTGCGCTCCGCGCTACCACTGCGCGCTTTCACGACAGGAGAAAGCTTGAAGTTC 1620
 Qy 1621 CTTGCGAAGCTGCGCTCAACAGAAAGCTCGGTAAGCTTCACTCTCTGAGAGAG 1680
 Db 1621 CTTGCGAAGCTGCGCTCAACAGAAAGCTCGGTAAGCTTCACTCTCTGAGAGAG 1680

Db 1681 CTTGCGAAGCTGCGCTCAACAGAAAGCTCGGTAAGCTTCACTCTCTGAGAGAG 1680
 Qy 1681 CTTGCGAAGCTGCGCTCAACAGAAAGCTCGGTAAGCTTCACTCTCTGAGAGAG 1680
 Db 1681 CTTGCGAAGCTGCGCTCAACAGAAAGCTCGGTAAGCTTCACTCTCTGAGAGAG 1680
 Qy 1741 GCGACCTTTCAGAGGAGCTGCAACAGAGAGCTCAAGAGCTCAAGATCAAGCTCGAAGAC 1800
 Db 1741 GCGACCTTTCAGAGGAGCTGCAACAGAGAGCTCAAGAGCTCAAGATCAAGCTCGAAGAC 1800
 Qy 1801 GCGCTCAAGACCATGAGAGGAGTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 1860
 Db 1801 GCGCTCAAGACCATGAGAGGAGTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 1860
 Qy 1861 GCGCTCAAGACCATGAGAGGAGTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 1920
 Db 1861 GCGCTCAAGACCATGAGAGGAGTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 1920
 Qy 1921 AAAGCACTTTACGTTGAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 1980
 Db 1921 AAAGCACTTTACGTTGAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 1980
 Qy 1981 CCGCGCTCGCTCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2040
 Db 1981 CCGCGCTCGCTCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2040
 Qy 2041 GAGGAGCGCTTTCGCTACAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2100
 Db 2041 GAGGAGCGCTTTCGCTACAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2100
 Qy 2101 GTTGAAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2160
 Db 2101 GTTGAAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2160
 Qy 2161 CAAATCGGTTTATGAGCTTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2220
 Db 2161 CAAATCGGTTTATGAGCTTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2220
 Qy 2221 GTTGAAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2280
 Db 2221 GTTGAAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2280
 Qy 2281 GCGACCAAGCTTTCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2340
 Db 2281 GCGACCAAGCTTTCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2340
 Qy 2341 TCCATCAGGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2400
 Db 2341 TCCATCAGGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2400
 Qy 2401 CAGAGAGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2460
 Db 2401 CAGAGAGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2460
 Qy 2461 GCGACCTTTCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2520
 Db 2461 GCGACCTTTCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2520
 Qy 2521 GCTGAGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2580
 Db 2521 GCTGAGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2580
 Qy 2581 GAGCGGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2640
 Db 2581 GAGCGGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2640
 Qy 2641 GAGCGGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2700
 Db 2641 GAGCGGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2700
 Qy 2701 GAGCGGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2760
 Db 2701 GAGCGGAGAGCTGCGCTCAAGAGAGCTCAAGAGAGCTCAAGAGAGCTCAAGATCAAGCTCGAAGAC 2760

QY 2761 CGCAGAACTTGGGGAGCCAAACCGACTGTGTGTGCTTTGGGAAAGACCGGCTAC 2820
| | | | |
Db 2761 CGCAGAACTTGGGGAGCCAAACCGACTGTGTGTGCTTTGGGAAAGACCGGCTAC 2820
QY 2821 GAGGTCTTTGGCGGTGTGCGAAATCAACGTAGACTTGGCGAACCCGACGGGACCCG 2880
| | | | |
Db 2821 GAGGTCTTTGGCGGTGTGCGAAATCAACGTAGACTTGGCGAACCCGACGGGACCCG 2880
QY 2881 AAGCCCATATAGGGCGTTTCCAGAAAGGGGTACAGTGGGTCAAGGTCAACACGGGTCAAC 2940
| | | | |
Db 2881 AAGCCCATATAGGGCGTTTCCAGAAAGGGGTACAGTGGGTCAAGGTCAACACGGGTCAAC 2940
QY 2941 AAACACAGGCGCTCCAGACGCTGTGTGTCCGCTTACCAAGCGAAAGCGTGAACCTGGCCG 3000
| | | | |
Db 2941 AAACACAGGCGCTCCAGACGCTGTGTGTCCGCTTACCAAGCGAAAGCGTGAACCTGGCCG 3000
QY 3001 CTACAGGAAGCTAAGAGAGAGCGTCAACCGCATGTCTAACTGCGTTGACCGACATTTGGGAC 3060
| | | | |
Db 3001 CTACAGGAAGCTAAGAGAGAGCGTCAACCGCATGTCTAACTGCGTTGACCGACATTTGGGAC 3060
QY 3061 TGGACTGTCTAGTAAGACGCGCGTGAACGAGCTGTCTTTCAGAGACCGAGCTCAAGTTCAAC 3120
| | | | |
Db 3061 TGGACTGTCTAGTAAGACGCGCGTGAACGAGCTGTCTTTCAGAGACCGAGCTCAAGTTCAAC 3120
QY 3121 CAACGCGGCGGACCGGTGAAGACCTGTGAGACCGACGACCCCTTACATCCGTGACATA 3180
| | | | |
Db 3121 CAACGCGGCGGACCGGTGAAGACCTGTGAGACCGACGACCCCTTACATCCGTGACATA 3180
QY 3181 GACTTCCTTATGAAGACTGACGAGAAAGTGTCCGCAAGCGCATCAATAGGGCAAGGTC 3240
| | | | |
Db 3181 GACTTCCTTATGAAGACTGACGAGAAAGTGTCCGCAAGCGCATCAATAGGGCAAGGTC 3240
QY 3241 GGGCAGGGAGTCCGCGCTCACTCAAAAGTCTCTCACTTCCGCTCCGCGTTGGATACGC 3300
| | | | |
Db 3241 GGGCAGGGAGTCCGCGCTCACTCAAAAGTCTCTCACTTCCGCTCCGCGTTGGATACGC 3300
QY 3301 ATACTCGAGAGATCTCCGTACCGGAGCGGACCGGTCCGGTACAGCAACGGGTCTCCG 3360
| | | | |
Db 3301 ATACTCGAGAGATCTCCGTACCGGAGCGGACCGGTCCGGTACAGCAACGGGTCTCCG 3360
QY 3361 GAGGAAGAAAGCGCATGCTGCTGAGAGCGAAGTCAATGATGCCACACGCGACGTTTC 3420
| | | | |
Db 3361 GAGGAAGAAAGCGCATGCTGCTGAGAGCGAAGTCAATGATGCCACACGCGACGTTTC 3420
QY 3421 GTCTCGGCGGACTGAGACCGAGTTTGAACCGGCCCAATTAACAGAGTGTCTCTTC 3480
| | | | |
Db 3421 GTCTCGGCGGACTGAGACCGAGTTTGAACCGGCCCAATTAACAGAGTGTCTCTTC 3480
QY 3481 GCGGCGCTTTTGAAGCGCATCGGACGCGCTGACGCTGCTTAATCTATTCAAGAAACGG 3540
| | | | |
Db 3481 GCGGCGCTTTTGAAGCGCATCGGACGCGCTGACGCTGCTTAATCTATTCAAGAAACGG 3540
QY 3541 TGGGGAAGAAAGCACTTGGAGAGGAGGTCTAGGCTCCGTTGAAGTGAAGCGGTCTGCTC 3600
| | | | |
Db 3541 TGGGGAAGAAAGCACTTGGAGAGGAGGTCTAGGCTCCGTTGAAGTGAAGCGGTCTGCTC 3600
QY 3601 GACTCCGCGCAGCTTGGAGCGCTTGGCGAACAACATCTTCTGCGCGCGTCAATGCTC 3660
| | | | |
Db 3601 GACTCCGCGCAGCTTGGAGCGCTTGGCGAACAACATCTTCTGCGCGCGTCAATGCTC 3660
QY 3661 ACGCTCTTCCGCGCGCTGCAAGGTTGCGAGCTTTCAAAAGCGACGACTGCTCTGTGGT 3720
| | | | |
Db 3661 ACGCTCTTCCGCGCGCTGCAAGGTTGCGAGCTTTCAAAAGCGACGACTGCTCTGTGGT 3720
QY 3721 AGCCATTACCTCGCTTTCGAGCGCTTGAAGCGCTTCAATGAGCGAAGCTTACAAAGCAAA 3780
| | | | |
Db 3721 AGCCATTACCTCGCTTTCGAGCGCTTGAAGCGCTTCAATGAGCGAAGCTTACAAAGCAAA 3780
QY 3781 CATTTGAAGGTGAGGTGAGAAATCGTGCCTGATCAATCGGACTCTGCTCCGCTGAG 3840
| | | | |
Db 3781 CATTTGAAGGTGAGGTGAGAAATCGTGCCTGATCAATCGGACTCTGCTCCGCTGAG 3840

QY 3841 CAGGTCTCTGAGACCTGTGTGAGAGCGCTCTCAAGATATTGGGCGTGTACACAAAC 3900
| | | | |
Db 3841 CAGGTCTCTGAGACCTGTGTGAGAGCGCTCTCAAGATATTGGGCGTGTACACAAAC 3900
QY 3901 GAACTCTCTTATCTCAAGATCTGAGAGGCTGTGAGAGATGATCAACAAAGGCTGAGTAC 3960
| | | | |
Db 3901 GAACTCTCTTATCTCAAGATCTGAGAGGCTGTGAGAGATGATCAACAAAGGCTGAGTAC 3960
QY 3961 GCCCGCTACCAAGCTCTGTGTGCAATGTGACAGATGTCTACTAATTAAGCGCGGAG 4020
| | | | |
Db 3961 GCCCGCTACCAAGCTCTGTGTGCAATGTGACAGATGTCTACTAATTAAGCGCGGAG 4020
QY 4021 TCTGCGGCTATATATGAGCGCTGTGTGTCTTGTGGCGGCGGACCTTCCGTTTGA 4080
| | | | |
Db 4021 TCTGCGGCTATATATGAGCGCTGTGTGTCTTGTGGCGGCGGACCTTCCGTTTGA 4080
QY 4081 CAATCGCGGTGTGCTGCTGATGTGAGAGACCGGACCGCTTACAGACAGATATCCG 4140
| | | | |
Db 4081 CAATCGCGGTGTGCTGCTGATGTGAGAGACCGGACCGCTTACAGACAGATATCCG 4140
QY 4141 GCTAACGTGCGGATGTGTGCTTGAACAGCTTTCAGACCCCGCGGACCGCGCCG 4200
| | | | |
Db 4141 GCTAACGTGCGGATGTGTGCTTGAACAGCTTTCAGACCCCGCGGACCGCGCCG 4200
QY 4201 GCGAGTTTGTGCGGACATGTGCGAAGCGGAAACGCTTCTTCACTTACCGGAAAGCT 4260
| | | | |
Db 4201 GCGAGTTTGTGCGGACATGTGCGAAGCGGAAACGCTTCTTCACTTACCGGAAAGCT 4260
QY 4261 GGTGTTTTCGCGACTCAAGCCAGCTTGGAGCTGAGGACCGGCGGAGTCTCATGG 4320
| | | | |
Db 4261 GGTGTTTTCGCGACTCAAGCCAGCTTGGAGCTGAGGACCGGCGGAGTCTCATGG 4320
QY 4321 GATGCACTGTGAGCAACAGCTTTTGGAGTTTGAACCGGAGACCCGCTCAATCA 4380
| | | | |
Db 4321 GATGCACTGTGAGCAACAGCTTTTGGAGTTTGAACCGGAGACCCGCTCAATCA 4380
QY 4381 TCTGTGCGGTGATGTGTCTTCAATGAGCTCTCTCAATGAGTGTGAGAGTGTCTCAATG 4440
| | | | |
Db 4381 TCTGTGCGGTGATGTGTCTTCAATGAGCTCTCTCAATGAGTGTGAGAGTGTCTCAATG 4440
QY 4441 GAGAGACCGGCAAGAGCAACAGAGACTTGAACAGAGAAAGCGGCTTGCAGACAG 4500
| | | | |
Db 4441 GAGAGACCGGCAAGAGCAACAGAGACTTGAACAGAGAAAGCGGCTTGCAGACAG 4500
QY 4501 CAATCAAGCTGTGTGAATGTCTGACAGAAAGGAGAAAGACAGGAGTTCTTAAT 4560
| | | | |
Db 4501 CAATCAAGCTGTGTGAATGTCTGACAGAAAGGAGAAAGACAGGAGTTCTTAAT 4560
QY 4561 GCCCGCACTGCTCCGAGCCCTCAATCTCAATTTTGGAAAGAGTGTGAGACCGG 4620
| | | | |
Db 4561 GCCCGCACTGCTCCGAGCCCTCAATCTCAATTTTGGAAAGAGTGTGAGACCGG 4620
QY 4621 CCGACTGTGCGCGCTGACATCACTTGGCAACCCCACTCTGCGCAACAGATT 4680
| | | | |
Db 4621 CCGACTGTGCGCGCTGACATCACTTGGCAACCCCACTCTGCGCAACAGATT 4680
QY 4681 GCCGCGAGAGCAACGCGGAACTTTGGCGCTTCTTGGGTTTCCAGTGTGCGGTGCTG 4740
| | | | |
Db 4681 GCCGCGAGAGCAACGCGGAACTTTGGCGCTTCTTGGGTTTCCAGTGTGCGGTGCTG 4740
QY 4741 TCGAAGCGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4800
| | | | |
Db 4741 TCGAAGCGGTACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 4800
QY 4801 TGGAAAGGCGTGAACCGGAGACCGAGAGTCAAGGAGAGACCGGAGACAGGCGCGGTC 4860
| | | | |
Db 4801 TGGAAAGGCGTGAACCGGAGACCGAGAGTCAAGGAGAGACCGGAGACAGGCGCGGTC 4860
QY 4861 GTCCAGGCTCTGATCAGCGGCGGCTTATCTTCAAGAAAGCAAGCTTCTTCCAGCATCC 4920
| | | | |
Db 4861 GTCCAGGCTCTGATCAGCGGCGGCTTATCTTCAAGAAAGCAAGCTTCTTCCAGCATCC 4920
QY 4921 AAAGGCTACTCAAGAACTAAGGAGTGTCAATTCACCTTTTCTG---CCCGAGT 4977

Db	4921	AAAGGCTTACTCAAGACATTAAGGAGTGTCTCACAATCTCACTCTTTTCTCGCTTCCCCAGT	4980
QY	4978	GCGGATTACCAAGGCCCCGCACTGCCAGACAGTCCGAGTCTG-----CGCGCCGCTGC	5030
Db	4981	GCGGATTACCAAGGCCCCGCACTGCCAGACAGTCCGAGTCTGCGGAGACCGCCGCTGC	5040
QY	5031	AGAAATGGCGCGCTCATGTATTATTCACAGACCGTTGGC-----TTCAT	5071
Db	5041	AGAAATGGCGCGCTCATGTATTATTCAGACCGCTTGGCAGACAGAAATTAGAAATTTCAT	5100
QY	5072	CTGCGCGACAGGCCGACTTGAAGGCATACGCTCACTCGGA-----CTCT	5118
Db	5101	CTGCGCGACAGGCCGACTTGAAGGCATACGCTCACTCGGACAGCGAGAAATTGCTCT	5160
QY	5119	GTTTCCGATGTAAAGATCAGCAGACGCATGAAGAACAAAT-----TA	5164
Db	5161	GTTTCCGATGTAAAGATCAGCAGACGCATGAAGAACAAAT-----TA	5220
QY	5165	GTTTCTTTGTTGCTAAACAGGTGGTCCCTCCATTTGAGGTAAAGACTCTGTGAGTCT	5224
Db	5221	GTTTCTTTGTTGCTAAACAGGTGGTCCCTCCATTTGAGGTAAAGACTCTGTGAGTCT	5280
QY	5225	CAACGTTTACTGTTGAGTCTGCTGGGTGCTTCATTCCCAACAGCAGAAAGGTGTGC	5284
Db	5281	CAACGTTTACTGTTGAGTCTGCTGGGTGCTTCATTCCCAACAGCAGAAAGGTGTGC	5340
QY	5285	AACTAGTACGGCGCCCTCGGGATTACCA	5312
Db	5341	AACTAGTACGGCGCCCTCGGGATTACCA	5368

RESULT 5
 US-09-991-262-47
 Sequence 47, Application US/09991262
 Publication No. US20030041349A1
 GENERAL INFORMATION:
 APPLICANT: Chritcian, P. D., Gordon, K. H. J., Hanzlik, T. N.
 TITLE OF INVENTION: Insect Viruses and Their Uses in
 Protecting Plants
 NUMBER OF SEQUENCES: 57
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Flehr Hobdach Test Albritton & Herbert LLP
 STREET: Four Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/991,262
 FILING DATE: 20-NO. US20030041349A1-2001
 CLASSIFICATION: <unknown>
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US/09/234,238
 FILING DATE: 20-JAN-1999
 APPLICATION NUMBER: US 08/485,355
 FILING DATE: 07-JUN-1995
 APPLICATION NUMBER: US 08/440,522
 FILING DATE: 12-MAY-1995
 APPLICATION NUMBER: US 08/089,372
 FILING DATE: 08-JUL-1993
 APPLICATION NUMBER: AU PL4081/92
 FILING DATE: 14-AUG-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Treccartin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-58631-3/RFT/DSS
 TELECOMMUNICATION INFORMATION:

INFORMATION FOR SEQ ID NO: 49:
SEQUENCE CHARACTERISTICS:
LENGTH: 2478 base pairs
TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 366..2306
SEQUENCE DESCRIPTION: SEQ ID NO: 49:
US-09-991-262-49

Query Match 1.3%; Score 71.2; DB 11; Length 2478;
Best Local Similarity 78.7%; Pred. No. 1.6e-09;
Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 5205 TAAAGACTGTGAGTCTCAAGCTTACTGTTGAGTCTGTCGGTTCGATTCATTC 5264
DB 2371 TAAAGACTGTGAGTCTCAAGCTTACTGTTGAGTCTGTCGGTTCGATTCATTC 2430

QY 5265 CCAAGCAGCAAGAGGCTGCGCACTAGTACGCGCCCTCGGATACCA 5312
DB 2431 CCAAGCAGCAAGAGGCTGCGCACTAGTACGCGCCCTCGGATACCA 2478

RESULT 7
US-09-991-262-51
Sequence 51, Application US/09991262
Publication No. US20030041349A1
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hobach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/09/991,262
APPLICATION DATA:
FILING DATE: 20-NO. US20030041349A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/234,238
FILING DATE: 20-JAN-1999
APPLICATION NUMBER: US 08/485,355
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Trecartin, Richard F.
REGISTRATION/DOCKET NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-3/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 2479 base pairs

TYPE: nucleic acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: DNA
FEATURE:
NAME/KEY: CDS
LOCATION: 283..2307
SEQUENCE DESCRIPTION: SEQ ID NO: 51:
US-09-991-262-51

Query Match 1.3%; Score 71.2; DB 11; Length 2479;
Best Local Similarity 78.7%; Pred. No. 1.6e-09;
Matches 85; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

QY 5205 TAAAGACTGTGAGTCTCAAGCTTACTGTTGAGTCTGTCGGTTCGATTCATTC 5264
DB 2372 TAAAGACTGTGAGTCTCAAGCTTACTGTTGAGTCTGTCGGTTCGATTCATTC 2431

QY 5265 CCAAGCAGCAAGAGGCTGCGCACTAGTACGCGCCCTCGGATACCA 5312
DB 2432 CCAAGCAGCAAGAGGCTGCGCACTAGTACGCGCCCTCGGATACCA 2479

RESULT 8
US-10-156-761-3243
Sequence 3243, Application US/10156761
Publication No. US20030119018A1
GENERAL INFORMATION:
APPLICANT: OMURA, SATOSHI
APPLICANT: IKEDA, HARUO
APPLICANT: ISHIKAWA, JUN
APPLICANT: HORIKAWA, HIROSHI
APPLICANT: SHIBA, TADAYOSHI
APPLICANT: SAKAKI, YOSHIYUKI
APPLICANT: HATTORI, MASAHIRA
TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
FILE REFERENCE: 249-262
CURRENT APPLICATION NUMBER: US/10/156,761
CURRENT FILING DATE: 2002-05-29
PRIOR APPLICATION NUMBER: JP 2001-204089
PRIOR FILING DATE: 2001-05-30
PRIOR APPLICATION NUMBER: JP 2001-272697
PRIOR FILING DATE: 2001-08-02
NUMBER OF SEQ ID NOS: 15109
SEQ ID NO 3243
LENGTH: 627
TYPE: DNA
ORGANISM: Streptomyces avermitilis
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(627)
US-10-156-761-3243

Query Match 1.0%; Score 51.6; DB 15; Length 627;
Best Local Similarity 52.9%; Pred. No. 0.0004;
Matches 111; Conservative 0; Mismatches 99; Indels 0; Gaps 0;

QY 1714 ACCGCGGCTCTACTCCCGAGTACGCGGCACCTTGACGAGCTTCAACGACGTC 1773
DB 304 AACCCGCGTCCCGCTCAACATCTACGCCACTTCAAGGCGCGAGCAAGGCAAG 363

QY 1774 AACGAGTCAAGATCACTCCCTGAGAAAGCCCTTCAACATGACGAGCTCACTCTCC 1833
DB 364 GACCAAGTCAAGGTATGCTGAGAAAGCCCGGTCTTGAAGTCTGCAAGCTCAAGCC 423

QY 1834 CCAAGCAGAGGCTCTGAGATGTATGAGAGGCGCCGACGAGCAAGAGGCAAGGCACTTC 1893
DB 424 CTGAGCCCGGCGGCAAGAGCAAGCAAGCCGCGGCAAGCCGAGAGGAGGCTCCCATC 483

QY 1894 ATGCGCGCTTGAAGCGCGGCGGTAAT 1923
DB 484 ACCTTGCGCTGACAGCGGCGGACCCAA 513

RESULT 9
US-10-156-761-1/c
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match 1.0%; Score 51.6; DB 15; Length 9025608;
Best Local Similarity 52.9%; Pred. No. 0.012;
Matches 111; Conservative 0; Mismatches 99; Indels 0; Gaps 0;
QY 1714 ACCGGGCTCTACCTCCCGAGTACCGCGCCACCTTGACGAGGAGCTCTCAACACCGACCTC 1773
DB 4052336 AACCCGGGCTCCCGGTCAACATCTACCGCACCTTCAAGCGCGCCGAGACCAAGGCGAAG 4052277
QY 1774 AAGCAGCTCAAGATCAACCTCGAGAACCCCTCAAGACCATCGACGAGCTCAACCTCTCC 1833
DB 4052226 GACCACTCAAGTATGTGTGAGAACCCCGCGTCTCGAGTGGCAAGTCAACCGCC 4052217
QY 1834 CCACTCAGAGGCTCTGAGATGTATCAGAGGCGCCGACGAGCGGACAGACGCGCACCTC 1893
DB 4052216 CTCGACCCCGGACAGAGGAGCAGCGCGCCGACCGGAGGAGGCGCTCCCATC 4052157
QY 1894 ATCCCGCGCTTGAGGCGCGCGCGGTAAA 1923
DB 4052156 ACCTTCGGCTCGACACGCGGACGCCCA 4052127

RESULT 10
US-10-156-761-6555
; Sequence 6555, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02

NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 6555
; LENGTH: 1650
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1650)
US-10-156-761-6555

Query Match 1.0%; Score 51.2; DB 15; Length 1650;
Best Local Similarity 52.9%; Pred. No. 0.0074;
Matches 110; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 1701 CTGTGACGCGACCGCGGCTCTACTTCCCGAGTACCGCGCCACCTTGACGAGGAGTCT 1760
DB 378 CTCGCCCCAGCGGCGCGCGCATCTGCGCGGAGGTCTCGCCCTGCGCGCGCGCA 437
QY 1761 CAACACCGAGCTCAAGAGCTCAAGATCAACCTTGAGAGAGCCCTCAAGACATCGACG 1820
DB 438 CACGCGGAGGTGTGACCGCGAGCTCAAGCCCGGCAACTCTGATGGGAAACGAG 497
QY 1821 GCTCACCTCTCCCGAGTCAAGGCTCTGAGATGTATCAGAGGCGCGCGACGCGCAA 1880
DB 498 CCGGAGGTGTCTACCGACTTGGCATCGCACGCTGAGGCGAGCTCGGCTGACCAT 557
QY 1881 GACGCGACCTCTATCGCGCGCTTGAG 1908
DB 558 GACCGGAGGTCTATCGGCTGCGCGGAG 585

RESULT 11
US-10-369-493-42399/c
; Sequence 42399, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 42399
; LENGTH: 918
; TYPE: DNA
; ORGANISM: Halobacterium sp. NRC-1
US-10-369-493-42399

Query Match 0.9%; Score 49.8; DB 12; Length 918;
Best Local Similarity 49.8%; Pred. No. 0.0015;
Matches 126; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

QY 666 CAAGTGTTAATGCCGAGCTCGGCTACCGCTTCCAGATTATGAAGGCTATGGCTGT 725
DB 295 CGAATGCTCTTCCCGCGCTCAAGAGAGTGTGTGAATTCAGAGGCGCGGAGTT 236
QY 726 GAAGACTCGCATTCAGAGGAGGAGACTTCGCTCCATCTCCCTGAGTTGACTTCAT 785
DB 235 CGTGAAGGCGGCGTCAACGAGGAGTCCGCGGACATCAGCAAGATGGGCTTCTCG 176
QY 786 CAAGAGAGCCAAAGAGCGGCGCATCGAGGCTGAGCGGCGCGGCTCTACTCAGAG 845
DB 175 CACGAGGCTCGCGGAGAGTCCGTCAGCGGTGTGCGGAGGTGCGCGCGCTCGCGCG 116
QY 846 GCGGTATTTTTCGCGGAGAGAGAGTGGGAGTATGCGATTTACAGCACTTCCACAG 905

```

Db      115 CGTCTCGGCGCGGCTGATGTGCAAGCGACGACGTCGTGGCCGCGCTGAGAC 56
Qy      906 ATGGCTGGCTTAC 918
Db      55 AGCGGTGCGACGAC 43

RESULT 12
US-10-156-761-4793
; Sequence 4793, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 4793
; LENGTH: 1449
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1449)
US-10-156-761-4793

Query Match      0.9%; Score 49.8; DB 15; Length 1449;
Best Local Similarity 48.4%; Pred. No. 0.0018;
Matches 138; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy      3093 TGTCTTGAGACCGCTCAAGTTTCAACCCGAAGCGGGGCGACCGTGAAGACCTGTGGA 3152
Db      990 TTTCTTCTGCTCCCTGTTGAAGATCATCTGCGGACATGTGAGACTACCACTTGCCCGA 1049

Qy      3153 GCCAGACGACCCCTACATCCGTGACATAGACTTCTTATGAGAATCAGAGAAAGTGTG 3212
Db      1050 GGGGGGGGCTTCCACAACTGGCGGATCGTCTCATGACAGAAAGTACCGAGACGCG 1109

Qy      3213 GCCCAAGCCGATCAATACGGGCAAGTGGGCGAGGGGATCGCGCTCACTCAAAATCTCT 3272
Db      1110 CCAGAAAGTGTATCACCGGCTGTGGGGGCGGCAATGATGTCTCTGACCAAGCTGATCGT 1169

Qy      3273 CAACCTGCTGCTGGCGCTTGATAGCATATCTCGAGAGATCTCCGTAACCGGAGCGG 3332
Db      1170 GATGCTGATCTCGACGTCGACGATCTGCAAGAGTGCCTTGCGGGGCGCTCGG 1229

Qy      3333 CACGGTCCGGTACAGCAACGCTCTCCCGGACGAGAGAGAGCCAT 3377
Db      1230 CAACAGGACTACGCGCCGCGACCTCAAGTGTGTGAAGGCCCGT 1274

RESULT 13
US-10-156-761-1
; Sequence 1, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI

```

```

; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 1
; LENGTH: 9025608
; TYPE: DNA
; ORGANISM: Streptomyces avermitilis
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: (4187715)
; OTHER INFORMATION: a, t, c, g, other or unknown
US-10-156-761-1

Query Match      0.9%; Score 49.8; DB 15; Length 9025608;
Best Local Similarity 48.4%; Pred. No. 0.04;
Matches 138; Conservative 0; Mismatches 147; Indels 0; Gaps 0;

Qy      3093 TGTCTTGAGACCGCTCAAGTTTCAACCCGAAGCGGGGCGACCGTGAAGACCTGTGGA 3152
Db      5851163 TTTCTTCTGCTCCCTGTTGAAGATCATCTGCGGACATGTGAGACTACCACTTGCCCGA 5851222

Qy      3153 GCCAGACGACCCCTACATCCGTGACATAGACTTCTTATGAGAATCAGAGAAAGTGTG 3212
Db      5851223 GGGGGGGGCTTCCACAACTGGCGGATCGTCTCATGACAGAAAGTACCGAGACGCG 5851282

Qy      3213 GCCCAAGCCGATCAATACGGGCAAGTGGGCGAGGGGATCGCGCTCACTCAAAATCTCT 3272
Db      5851283 CCAGAAAGTGTATCACCGGCTGTGGGGGCGGCAATGATGTCTCTGACCAAGCTGATCGT 5851342

Qy      3273 CAACCTGCTGCTGGCGCTTGATAGCATATCTCGAGAGATCTCCGTAACCGGAGCGG 3332
Db      5851343 GATGCTGATCTCGACGTCGACGATCTGCAAGAGTGCCTTGCGGGGCGCTCGG 5851402

Qy      3333 CACGGTCCGGTACAGCAACGCTCTCCCGGACGAGAGAGCCAT 3377
Db      5851403 CAACAGGACTACGCGCCGCGACCTCAAGTGTGTGAAGGCCCGT 5851447

```

```

RESULT 14
US-10-369-43237
; Sequence 43237, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 43237
; LENGTH: 1467
; TYPE: DNA
; ORGANISM: Myxococcus xanthus
US-10-369-43237

Query Match      0.9%; Score 49.6; DB 12; Length 1467;
Best Local Similarity 45.7%; Pred. No. 0.002;
Matches 172; Conservative 0; Mismatches 204; Indels 0; Gaps 0;

```

QY 1483 CGCGTCGTGATAGTACCGAGTCCACTCGCTCGAGACATATCTGCGACGTCGCGCTG 1542
 DB 757 CGCGACGTGAGAGCGCGCGACGTGCTCTCTTCTGAGACATCTTCCGCTTACCCAG 816
 QY 1543 TCCCGTGAACAGATCGGCTTCTGCGCTCGCGGTACACCTGCCGCGCTTTCAGAC 1602
 DB 817 GCGGCGTCGAGAGTGTCCGCTCTGAGCGCGATCCAGCGCGGTGAGTTACAGAGCC 876
 QY 1603 AGGGAAGAGCTTGAAGTCTTGGCGAAGCTGGCTGTACAGAGACGTCCGTAACCTTC 1662
 DB 877 AGCTGGCCACGAGATGGCGGTCTGCAAGAGCGCATCCTCCACAGAGAGTTCC 936
 QY 1663 ACTCCTCGTGGAGAGCGCCCAAGTTTGAAGCGCCACTTGTGAGCGGACCGCGGCC 1722
 DB 937 ATCACTCTCGTGAAGCGCATCTACGTCCCGCGACGACCTCAAGAGCCGGGCGCTCGC 996
 QY 1723 TCACTCCCGAGTACCGCGCACCTTGCAGAGAGTCTCAACACGAGCTCAAGAGCTC 1782
 DB 997 AGGCGCTTGGCCACCTCGACGAGAGAGGTGCTCAACGCGCTCATCGCGAGCTCGCC 1056
 QY 1783 AAGATCACCTCTGAGAGCGCCTCTAAGACCATGACGAGCTCACTTCCCAATGAGA 1842
 DB 1057 ATCTTCCCGCGGTGAGCCGCTGAGACTCCACGAGCGCATCTGAGACCGCGCGTATC 1116
 QY 1843 GGCCTCGAGATGTACG 1858
 DB 1117 GCGCGGAGGACTACG 1132

RESULT 15

US-10-156-761-4716
 ; Sequence 4716, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMIURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 4716
 ; LENGTH: 1047
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1047)
 US-10-156-761-4716

Query Match 0.9%; Score 49; DB 15; Length 1047;
 Best Local Similarity 51.1%; Pred. No. 0.0027;
 Matches 115; Conservative 0; Mismatches 110; Indels 0; Gaps 0;

QY 1694 AGCGCAGTCTGAGCAGGACCGCGCTCACTCCCGAGTACCGCGCCACTTGCAGG 1753
 DB 68 AGGTGACCGGGATGCGCGTGCAGCGGTCTGCGCGATCATCTGGGCGGTGCGG 127
 QY 1754 CAGGTCTCAACCGAGCTCAAGAGCTCAAGTCAACCTCTGAGAAAGCCCTCAAGACCA 1813
 DB 128 ACTTCATCCAGGCGCATTAAGGACGCCACATCGGCGCGTCTGCGCGAGCACTTCG 187
 QY 1814 TCAGACGGGCTCACCTCTCTCCCGATCAAGAGGCTCTGAGATGTAGAGGGCCGCAAGCA 1873

DB 188 TCAGACAGGTGATCCATGAGACGTGACCGGACCGCGCTGAGGACCGGAGCGGCA 247
 QY 1874 GGGCAAGACGGGCACTCTCATTCGCGCTTGAAGCGCGCGGCG 1918
 DB 248 CGGTCAAGGAGACCAACGTCAATCGGACCAATGACAGCTGCTGGGTG 292

Search completed: January 15, 2004, 16:47:28
 Job time : 1231 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 13:16:11 ; Search time 63.0574 Seconds
(without alignments)
4289.266 Million cell updates/sec

Title: US-09-991-262-40

Perfect score: 8832

Sequence: 1 MYKATIVAAVYAAADVAYV.....LKRIRSTSDSVPVKISKSA 1704

Scoring table: BLOSUM62

Gapop 10.0 , Gapept 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8784	99.5	1704	15	AA49657
2	1444.5	16.4	1233	19	AA41935
3	1443.5	16.3	1233	19	AA43536
4	520	5.9	1707	22	AA678760
5	515	5.8	1698	20	AAV31381
6	512.5	5.8	1708	20	AAV31384
7	510.5	5.8	1708	20	AAW93408
8	506	5.7	1693	19	AAW76368
9	500	5.7	1693	15	AAW51264

10	500	5.7	1693	19	AAW61519	Hepatitis B virus
11	497	5.6	1693	17	AA91813	Hepatitis B virus
12	479	5.4	1693	19	AAW60196	Protein encoded by
13	479	5.4	1693	21	AAW24119	Hepatitis B virus
14	479	5.4	1693	22	AAW62522	HEV-Burma strain v
15	479	5.4	1693	23	AAO15658	Hepatitis B virus
16	474	5.4	1693	19	AAW71209	Protein encoded by
17	453	5.1	1693	12	AAW14618	Protein encoded by
18	314.5	3.6	1616	23	ABW04601	Tomato mosaic viru
19	294	3.3	1694	24	ABG76437	Sindbis plant viru
20	290	3.3	1116	15	AAW60608	Tobamovirus replic
21	284.5	3.2	1616	23	AAW18346	Tobacco mosaic vir
22	284.5	3.2	1616	23	AAW18348	TMV 183kDa mutant
23	284.5	3.2	1616	23	AAW19947	TMV 183kDa mutant
24	284.5	3.2	1616	23	AAW20138	TMV 183kDa mutant
25	284.5	3.2	1616	23	AAW20139	TMV 183kDa mutant
26	279	3.2	1116	23	ABW04603	Tomato mosaic viru
27	261	3.0	1116	19	AAW59276	Rubella virus RA27
28	259	2.9	1116	23	AAW18345	Tobacco mosaic vir
29	259	2.9	1116	23	AAW18347	TMV 126kDa mutant
30	259	2.9	1116	23	AAW19946	TMV 126kDa mutant
31	259	2.9	1116	23	AAW20136	TMV 126kDa mutant
32	259	2.9	1116	23	AAW20137	TMV 126kDa mutant
33	250.5	2.8	2205	16	AAW79048	Infectious rubella
34	246	2.8	1645	22	AAW31997	Amino acid sequenc
35	246	2.8	1645	23	ABG78960	CGMV isolate 4 18
36	241.5	2.7	1644	22	AAW32000	Amino acid sequenc
37	241.5	2.7	1644	23	ABG78963	CGMV isolate SH 1
38	236	2.7	487	23	ABJ04794	Hepatitis B virus
39	234.5	2.7	487	23	ABJ04788	Hepatitis B virus
40	231	2.6	2237	21	AAW58148	GRV-3 polypeptid
41	230	2.6	1144	22	AAW31995	Amino acid sequenc
42	230	2.6	1144	23	ABG78958	CGMV isolate 4 12
43	228.5	2.6	161	24	ABG76429	Brome mosaic virus
44	227.5	2.6	1143	22	AAW31998	Amino acid sequenc
45	227.5	2.6	1143	23	ABG78961	CGMV isolate SH 1

ALIGNMENTS

RESULT 1	AA49657	standard; Protein; 1704 AA.
ID	AA49657	
XX	AA49657	
AC	AA49657	
XX	25-MAR-2003	(updated)
DT	12-SEP-1994	(first entry)
DT	12-SEP-1994	
XX	Sequence of Heliothis armigera replicase encoded by RNA 1.	
DE	Sequence of Heliothis armigera replicase encoded by RNA 1.	
XX	HasV; RNA 1; small RNA virus; replicase.	
KW	HasV; RNA 1; small RNA virus; replicase.	
XX	Heliothis armigera stunt virus.	
OS	Heliothis armigera stunt virus.	
XX	WO9404660-A1.	
FN	WO9404660-A1.	
XX	03-MAR-1994.	
PD	03-MAR-1994.	
XX	13-AUG-1993;	93WO-AU00411.
PP	13-AUG-1993;	93WO-AU00411.
XX	14-AUG-1992;	92AU-0004081.
PR	08-JUL-1993;	93US-0089372.
PR	08-JUL-1993;	93US-0089372.
XX	(CSIR) COMMONWEALTH SCI & IND RES ORG.	
PA	(PACT) PACIFIC SEEDS PTY LTD.	
XX	Christian PD, Gordon KHJ, Hanzlik TN;	
PI	Christian PD, Gordon KHJ, Hanzlik TN;	
XX	WPI; 1994-083180/10.	
DR	WPI; 1994-083180/10.	
DR	N-PSDB; AA058522.	
XX	N-PSDB; AA058522.	

PT Small RNA virus capable of infecting insect species, e.g.
PT Heliothis - and transgenic plants contg. viral nucleic acid, for
PT protection against insect pests
XX
XX
XX Disclosure; Figure 1; 183pp; English.
CC The inventors claim a virus comprising a genome hybridizable with
CC the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are
CC those given in Figs 1 and 2 of the specification. Isolated proteins
CC or polypeptide prepn. of the proteins or polypeptides derivable
CC from the virus are also claimed.
CC H. armigera larvae were raised and viral RNA was extracted. The virus
CC RNAs were reverse transcribed into cDNA. Clone E3 represents 99.7%
CC of RNA 1. (hr236 contains about 88% or RNA 2.) The full length clone
CC of RNA 1 was completed using PCR. RNA 1 encodes a protein of mol.
CC wt. 187,000 which is regarded as the replicase in view of its AA
CC sequence similarity in certain limited regions to replicases of
CC of other RNA viruses. The apparent mol. wt. of this protein upon
CC in vitro translation of virus RNA and SDS-PAGE is 195,000. The
CC sequence given in Figure 1 is in the same sense as the viral
CC (positive-sense) RNA. There are other small ORFs at the 3' end,
CC corresp. to the proteins P1a, P1b and P14.
CC (Updated on 25-MAR-2003 to correct PN field.)
XX
XX Sequence 1704 AA;
SQ
Query Match 99.5%; Score 8784; DB 15; Length 1704;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1694; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 661 AILRRATLKGAPFATVVIDGCPMPPLVVAIVHALSPSSRYLVCDVHQIGIFIDQGTSA 720
DB 661 AILRRATLKGAPFATVVIDGCPMPPLVVAIVHALSPSSRYLVCDVHQIGIFIDQGTSA 720
QY 721 NMPILVADVKKCRRTFQTRCPADVYATFPQSLVYGCCTTSGCVASISGVADVYNS 780
DB 721 NMPILVADVKKCRRTFQTRCPADVYATFPQSLVYGCCTTSGCVASISGVADVYNS 780
QY 781 QAQTLCTQOEKSRHAGAMTVHEAQRTFASVILHNGSTABOKLAESHLLVGTTR 840
DB 781 QAQTLCTQOEKSRHAGAMTVHEAQRTFASVILHNGSTABOKLAESHLLVGTTR 840
QY 841 HTNHLIYRDPDGIERQLNHSKAKAEVFPDIPAPLEITTVKSEBYORNEWMAITPQSAT 900
DB 841 HTNHLIYRDPDGIERQLNHSKAKAEVFPDIPAPLEITTVKSEBYORNEWMAITPQSAT 900
QY 901 PHGATHLRKXNGDPDGCVALAATGYEVFGRAKINVELABPDATKPHRAPQEGVOW 960
DB 901 PHGATHLRKXNGDPDGCVALAATGYEVFGRAKINVELABPDATKPHRAPQEGVOW 960
QY 961 VKVTNASKHQALQTLISRYTKRSADLPHEAKEDVCKMNSLDHMDMTVEDARDAY 1020
DB 961 VKVTNASKHQALQTLISRYTKRSADLPHEAKEDVCKMNSLDHMDMTVEDARDAY 1020
QY 1021 FEYOLKFTQGGTVEDLLEPDDPYTRIDIPLMKTQOKVSPKPINTGKVGQIAHNSKSLN 1080
DB 1021 FEYOLKFTQGGTVEDLLEPDDPYTRIDIPLMKTQOKVSPKPINTGKVGQIAHNSKSLN 1080
QY 1081 FYLAAMIRLIEELITGSTRVYNSGLPDEEAMLEAKINOVPAATVSDMTFEDTAH 1140
DB 1081 FYLAAMIRLIEELITGSTRVYNSGLPDEEAMLEAKINOVPAATVSDMTFEDTAH 1140
QY 1141 NNTSELFAALLERIGTPAAAVNLFRERCGKRTLRKAGLSVYVDGLDLSGAMTPPCRT 1200
DB 1141 NNTSELFAALLERIGTPAAAVNLFRERCGKRTLRKAGLSVYVDGLDLSGAMTPPCRT 1200
QY 1201 ITSAAVMLTLFRGVYPAAFKGDLSLCSGSHYLRPDAISRLNGERYKTHLKYEVOKIYV 1260
DB 1201 ITSAAVMLTLFRGVYPAAFKGDLSLCSGSHYLRPDAISRLNGERYKTHLKYEVOKIYV 1260
QY 1261 IGLVASBOVYLDPRSAALKTIGRCYTSSELYSKVYEAVRDITKGMSPARVHSLCHNSA 1320
DB 1261 IGLVASBOVYLDPRSAALKTIGRCYTSSELYSKVYEAVRDITKGMSPARVHSLCHNSA 1320
QY 1321 CYNTNAPBSAAIYIDAVVAFGRGDPFBQLRVVRAHVQAPDAVSSSTYPANVPAASCLDHVF 1380
DB 1321 CYNTNAPBSAAIYIDAVVAFGRGDPFBQLRVVRAHVQAPDAVSSSTYPANVPAASCLDHVF 1380
QY 1381 EPRQAAAPGAVATCAKPTPSGLTAKAGVASTTSHVATGTAPPSBPMDDAPANPSSELL 1440
DB 1381 EPRQAAAPGAVATCAKPTPSGLTAKAGVASTTSHVATGTAPPSBPMDDAPANPSSELL 1440
QY 1441 TRETPTSSPSSSSSDSSTSGRSLSGDGTARTTEDLSRKPSPGSDRGRSBSCLDRSG 1500
DB 1441 TRETPTSSPSSSSSDSSTSGRSLSGDGTARTTEDLSRKPSPGSDRGRSBSCLDRSG 1500
QY 1501 BRTGSSLTAPTAAPSPSPSBRARLATGPTVAATSPSATPCATQOVAARTTPDEAPFL 1560
DB 1501 BRTGSSLTAPTAAPSPSPSBRARLATGPTVAATSPSATPCATQOVAARTTPDEAPFL 1560
QY 1561 GSGSARAVSKPYRPTTAAWKVTPPLHANKGYTGDBPBYREDBETAAYVQALISGAYPOK 1620
DB 1561 GSGSARAVSKPYRPTTAAWKVTPPLHANKGYTGDBPBYREDBETAAYVQALISGAYPOK 1620
QY 1621 TKLSDASAGYRTKSCGOSTPAPASADYQARDQTVAVCAAAEMARSCIHPLAASA 1680
DB 1621 TKLSDASAGYRTKSCGOSTPAPASADYQARDQTVAVCAAAEMARSCIHPLAASA 1680
QY 1681 ASADLRIRSTDSVDPVXISKA 1704
DB 1681 ASADLRIRSTDSVDPVXISKA 1704


```

QY 1180 GSEVVDGLDGSAAWTPCENITFSAAVMLTLFPGVKAFAFGDSSLICSHYLRFDA SRL 1239
DB 1123 GSDVSPFMDLSGAVMTIARTTLFASGLMLALFVGVDITAKGDVFLAGNN-LYLDABRL 1181
QY 1240 HMGERRTKLKYEVOKIVPIYIGILVSAEQVVDVPSALKIFGRCTYSEL 1291
DB 1182 RMGSYLAANNLKIEKTRAVSFGIPIVSGAAVTVADVRLARTGRSVKXNM 1233

RESULT 3
AAW34536
ID AAW34536 standard; Protein; 1233 AA.
AC AAW34536;
XX
XX 08-JUN-1998 (first entry).
XX
XX Nudarelia beta virus undefined protein.
XX
XX Vaccine; insecticide.
XX
XX Nudarelia beta virus.
XX
XX Key Location/Qualifiers
FH Misc-difference 1029
FT /note="encoded by GCN"
XX
XX MO9746666-A1.
XX
XX 11-DEC-1997.
XX
XX 02-JUN-1997; 97MO-AU00349.
XX
XX 31-MAY-1996; 96AU-0000234.
XX
XX (CSIR) COMMONWEALTH SCI & IND RES ORG.
XX
XX Gordon KH, Hanzlik TN;
XX
XX WPI; 1998-042175/04.
XX
XX N-PSDB; AAT99118.
XX
XX Modified small RNA viruses and virus-like particles - have altered
PT or substituted Ig-like domains to modify host cell tropism, useful
PS as insecticides and in medicinal applications
XX
XX Disclosure; Figure 2; 41pp; English.
XX
XX The sequence is that of an undefined protein translated from cDNA
CC encoding the viral genome RNA.
XX
XX Sequence 1233 AA;
XX
Query Match 16.3%; Score 1443.5; DB 19; Length 1233;
Best Local Similarity 31.0%; Pred. No. 2.1e-103; Indels 227; Gaps 44;
Matches 425; Conservative 195; Mismatches 525;
QY 5 ATDVAVYAAADVAAYAVLQRAVKLDFAPELKALETILHLYLRFKGTLPPTQPHIL 64
DB 4 ASKQRLVLDAGQERAKA-----AFQLDPIASVETLEDAQEKEGEMFRSGTKLPSTHKL 57
QY 65 AGHORVAEEVLHNFARGRSTYLEIGPSLHSLKILHGAENAFVADYHGTCTKKTGDSRH- 123
DB 58 AIDLRVAEKDLRRHVKNVPTVLEIGPSVESV--RYAVQTRDKERHGTCTSDADNLRN 115
QY 124 ---ITAESRSVATGRPEFKADSLANGIARTCEVQDVGVSAPKSRGVIGANSLSY--- 177
DB 116 KIGCEAHYDRKIG-----PDARLLAAGIPTDTFCVDGFSNCSYQSPALAIACHSLIPDG 168
QY 178 -DVTLEELANAFENHGLHMYRAFNHMEBELLYMDNVVNAELGYRPHVIEBMAVKDCAFQ 236
DB 169 ESNISIMVAKGMALHGHVIVAYAMHLLPVELLTITLDADNIFEGYSI----- 213

```

```

QY 237 GGDRLRHPPELDIPINSQRRRIERLARGSYSRAVIFAGDDMDGADLYLHFTWTLAYIL 296
DB 214 -----REBETGALCCTKRKAIFSGYNDPGSAVYHADAHMGWLK 253
QY 297 VRYNPPFGFSIAIEVQRHSGSIELRTTRAPGDHMLVVPRTSQGLCPINP--FYVA 354
DB 254 HRGVDTPYGFSLIDIQGRFGMHTKILIRGHSSSGISTVYPLSKQLMVPNIVIMYP 313
QY 355 DASGTEKTIILSQHVNMLNFMOTRPEK--ELVDMTILMSFAPARLALIVASVTES 412
DB 314 KAK-HEBEYIVTDKKYEBGVYVGTTRVQSGKSITLARIWQYIRRLTRIILNGTVHEK 372
QY 413 SNISPADIVRTVSLVYLIIRRRRAAVAVTKAXDVGESTFMSLKHVLSGCCGLRN 472
DB 373 TWTIASQDIERLAVSIFPKRQVRAVSEKALNRQKK-----CKSAB 414
QY 473 LKGTDVVFTKRVV---DKYR-----VHSLADYICDV-RLSP--EQ 506
DB 415 KQALLPVMNRRIANMFQDKFQIDEBVVRKYLECLKAQPMIHADKVYNGCTKRYNPVAE 474
QY 507 VGFPLSRVPPARVFDHRSBELVLRAGCTNERVPSSTPRVBEQGFADLMATAASLPS 566
DB 475 VG-----PKMHLATGTLRELORE-----IPSANBEPQDGAKAMSHAHADLDI 517
QY 567 YRATLQAGLNTD-----VKQKITLESNALKTIDGLTSPVRGLEM--YEGPPSGKT 616
DB 518 Y---ABGLRLDSEAKAAGKSLATLLOQAFQVLAKTKCEGNNIIEIYWTBPPSGKS 573
QY 617 GTLIAALEAAGKALVYAPTRBELREAMDRIKPPSASATQHVALLILRATAAGAFATV 676
DB 574 RAKPRPADLQGGVLCAPRTLRDLDSVNHPSVCTHNLNHYAAKSG--NRPRDVI 632
QY 677 VIDECEMPLVYTAIVHSLSPSSRIYLVGDHVOIGTIDFGCTANMPLVDVVKQCRRT 736
DB 633 VIDEASTTPACVYGTMHNASPPSSRIYLCGDPHGIYIDFSDRQDLKPFSSIIAECSTR 692
QY 737 FNOTKRCPADVAVTAFPOSLYPGCTTSGCVASISHVA---PDRYSQAOTLCTFOEKS 793
DB 693 FNTTYRCPODVNLPIFKTLYPDAISFSKQLSIRLYTRASVTRRHAATL--TDQKP 750
QY 794 RHGAGAMTVHAQGTTPASVILHNGSTAEQKLAESKSHLVGTRTNHLYIRPTGD 853
DB 751 H--SEPPVTHNQAR--TDVIVHAGTLEBRLLKVRNINVALRHTALYIRD---- 803
QY 854 IERQLNHSAAKAVFTDIPAPLEITTVK--PSEE-----VORNEVATIPQSATPHG 903
DB 804 -----ESEKGEIIVPSLMTPPSWSTYRCTVVDKVPDPAVEREN--GSSGCDSHHG 855
QY 904 AIIHLRNKPGDQDQCGVALAKTGYEFGRAKINVELABPDATPKRHR----- 952
DB 856 AITTLQE--LQKLTD-----TKGVRF-----BSEAVPTAHRVVLGDNDSC 896
QY 953 -----AFQ--EGVQVNVKVTNANSGQALQTLSSRYTKSADLPHEAKEDYKMLNSDRH 1006
DB 897 PDRIYMYQPTNLNGTKYTNIKDQALHTLVGARYKINS-----SSEETPSLTLRBSQPA 952
QY 1007 WDNTVTEADARAV---FETQLKFTQRGSTVEBLLERPDYIID--IDPLMKQOKY-S 1059
DB 953 QEWIIPRHAAPREVQDSCFATPCKRCERGHGVDDI---EDRWSNBGCRISYHLKQOQKYM 1009
QY 1060 PKPINTGKVGQGLAASKSLNPFVLAAMIRLSELRIGSTRVYNSGLPDBEAMLEAK 1119
DB 1010 PTKL---KLDQGISAEKCANIKLSAMVRITIDQOMSTSEKFT--FANQOSRDYMSITLAK 1065
QY 1120 INOVPAATFVSAMTBEEDTAHNTSBLFPAALERIGTPAAAVLFPERRCGKTLRAKGL 1179
DB 1066 L-QEKAREFKSIDIKEDVTANMVSVILVFSWRCDR--GCEHLLBYFEKRSKSTLSR-I 1122
QY 1180 GSEVVDGLDGSAAWTPCENITFSAAVMLTLFPGVKAFAFGDSSLICSHYLRFDA SRL 1239
DB 1123 GSDVSPFMDLSGAVMTIARTTLFASGLMLALFVGVDITAKGDVFLAGNN-LYLDABRL 1181
QY 1240 HMGERRTKLKYEVOKIVPIYIGILVSAEQVVDVPSALKIFGRCTYSEL 1291

```


Db 1323 LVGRVGRRT---KLVEAHTDVGRSLN---HFIPELGPIINTVTCB-----LYELVEAM 1369
 QY 1028 TORGGVIEDLLEPD--DPYTRDIDFLMKTOOK-VSRKPINTGVGGGIAHSHSLNFWLA 1084
 Db 1370 VERGGDGSANVLELDLCSRDVSRITFPQKCNKFTGTETIAHGVCIGISMSKTFALPRG 1429
 QY 1085 AMIRIDE-ELIRGSTRVRSNGLPDEEBAMLEAKINQVPHATFVSADTEPDTAMNNT 1143
 Db 1430 PMFRAIEKEITLALAPNVFYG---DAYEDTVLAANAAGAPGCKFENDSEFDSIQNNF 1485
 QY 1144 SELLPALALERICTPAAAVNLFRERCGKRTLRAGLGSVEVDGL-----LDGSAWTPGR 1198
 Db 1486 SLGLECIIMECEGPMQWIRLY-----HLVRSAMILQAPKESLRGPMWKHSGEPGTLIM 1539
 QY 1199 NTFSAAVMLTL--FRGVFPAARFGDS-LLCGSHYLRFPASRLHNGERYKTKLKYEVQ 1255
 Db 1540 NTVMNNAVIAHCYEFPLDKVAARFGDSVVLCSDYQSRDAVLAIGCGLK--LKYDFR 1596
 QY 1256 KIIPYIGLVSAEQVVL-DPVRSA 1278
 Db 1597 PIGLVAGVVAPGLGTLDPVVRPA 1620
 RESULT 5
 AAY31381
 ID AAY31381 standard; Protein; 1698 AA.
 AC AAY31381;
 DT 12-OCT-1999 (first entry)
 DE HEV-US1 ORF1 protein.
 XX Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
 KM vaccine; passive immunisation.
 XX Hepatitis E virus.
 OS
 XX
 PH Key Location/Qualifiers
 FT Misc-difference 174 /label= unknown
 FT /note= "encoded by ATR"
 FT Misc-difference 363 /label= unknown
 FT /note= "encoded by RTG"
 FT Misc-difference 1088 /label= unknown
 FT /note= "encoded by CRC"
 FT Misc-difference 1131 /label= unknown
 FT /note= "encoded by YTC"
 FT Misc-difference 1217 /label= unknown
 FT /note= "encoded by CRC"
 FT Misc-difference 1389 /label= unknown
 FT /note= "encoded by GWC"
 XX
 PN WO9919732-A1.
 PD 22-APR-1999.
 XX
 PF 15-OCT-1998; 98WO-US21941.
 PR 15-OCT-1997; 97US-0061199.
 PA (ABBO) ABBOTT LAB.
 PI Dawson GI, Deesai SM, Erker JC, Mushahwar IK, Schlauder GG,
 XX WPI; 1999-288017/24.
 DR N-PSDB; AA200195.
 XX

PT Detection of United States isolates of hepatitis B virus
 XX
 PS Claim 6; Page 175-180; 260pp; English.
 CC The invention provides a method for detecting a US (sub)type hepatitis B
 CC virus (US-HEV), or its naturally occurring variants in a sample by
 CC treatment with a binding partner specific for a marker of the virus, and
 CC then detecting any complex formed. The method is used to diagnose
 CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
 CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
 CC are useful in vaccines or for passive immunisation. The polypeptides are
 CC also used to raise specific antibodies (useful as immunoassay reagents).
 CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
 CC usual hybridisation and amplification assays for detecting infection. The
 CC present sequence represents a HEV-US1 ORF1 protein.
 XX
 SQ Sequence 1698 AA;
 Query Match 5.8%; Score 515; DB 20; Length 1698;
 Best Local Similarity 21.7%; Pred. No. 2.9e-30;
 Matches 397; Conservative 179; Mismatches 553; Indels 700; Gaps 84;
 QY 13 AAADVAVANVLQORAVKIDFAPPLKALET--LHRLYTP--LRPKGGLTPPTQHPILAGHQ 68
 Db 13 AAANSALANAVVVR-----PFLSRVQTEILINLMQRFQVFRPEVL--WNHPI----Q 59
 QY 69 RVAEBVLNHPARGRS-TVLRIGPSLSHSLKLGAPAPADYHGC-TKGTGDSRHITA 126
 Db 60 RVIHNELEQYCRARQCLEGAHPRS--INDNPV-----LHRCFLRVGSDVOQWYSA 112
 QY 127 LE-----SRVATGRPFKADASLALANGISRTFCVDGSCAPKSRVGIANHSLYOV 179
 Db 113 PTRGPANCRSRALRLP-----ADRTYCPGPGRCAPAAETGVALYSLHL 160
 QY 180 TLEELANAEENGLHNVRAFMKMPBELLYMDNVNNAELGYRPHVIEBPAAVDCAFQGD 239
 Db 161 WPADVABAMARHGXTLYAALHLPREVLPPGYHTTSLYLH-----DGD 206
 QY 240 LRHLPELDPFINESQRRIRBLAARSGYRRRAVIEGDDDKDAYLHDHTMALYL----- 295
 Db 207 -----RAVVTYEGDTSAG--YNHDSIILRAMVTRTK 235
 QY 296 LVENVPTPFGFSIHIEVORRHGSSIELRITRAP-----PDRLMLAVPRTSGLCRI 347
 Db 236 IVDHR-----LVIRVPAIGHFVLLTAAPBSPMPYVPRSTBYVASIFPGGS 289
 QY 348 PNIFYYADASGTEHKTILTSOHKVNMLNFMQTRPEKBLVDMTVMSFARARLAIIVAS 407
 Db 290 PSLF--PSACSTKSTHAPVPHIMDRMLFGATLDQARC-CGRMLTYLRGISYKTVGA 346
 QY 408 EYTESSWNISPADLVRTVVSIVYLHIER-----RRAVA----- 442
 Db 347 LVANEGMNABEDLTXXITAAAYLTICORLYRTQALSKGMRRLGVHAOKFITRLYSWF 406
 QY 443 VKTAKDDV-----FGTSFWS-----LKHVLGS--CGGLR 471
 Db 407 EKSGRDIYERQQLFYAQCRRMLSAGFLHDPRLVLDVDSVPCRCRTFLKKAAGKCCFMR 466
 QY 472 NL-----KGTDV----- 478
 Db 467 WLQGECTCFLEPABGLVGHDHNDNAVBESGVDPAPAPHLDVSGTYAVHGHQLEALYRAL 526
 QY 479 -----VTKRVVDKRV----- 490
 Db 527 NVPQDIAARASRLTATVELVASEPDLRCRTVLGKTKFTTVVDGAHLBANGPQOYLSPD 586
 QY 491 -----HSL-----GDI--ICD----- 499
 Db 587 ASRQSGAGSHSLTYELTRAGLOVRISSNGDCTATPPGGAASAPBGRVAACSLXYR 646
 QY 500 -----VRLSPEOV-----GF----- 509
 Db 647 NREFGHSLTGLMLHPEGLGIFPPFSQHIWESANPFCGEGTLVTRTWSTSGFSSDFS 706

```

QY 510 -----LPSKVPAR-----VFHRELEVLREAGCYNRPVST----- 543
Db 707 PPEAABAMATPGLPHSTPPVSDIWWLPPSEEFQV-----DAAPAPADPAGLPG 759
QY 544 -----PVEBP-----OGFDADL-W----- 557
Db 760 PVULTPPPPVVKHPSIFPPSRNRLLYTPDGAKVYAGSLFESDCMLVNASPNRHRPG 819
QY 558 -----HATAASLPEYRA-----TLQAG- 574
Db 820 GGLCHAFYQRPPEAFYPTFPIRBGLAAYTLTPRIHAVA-----PDYVVEGNPKRLBAY 876
QY 575 -----LNDVQKQKITL----- 586
Db 877 RETCSRRTAAVPLGSGIYQVPVSLSPAWERNHRPGDELYTEPAAMWEANKPAQPV 936
QY 587 -----ENMLKT-----ID-----GLTSPVSGLEMYE-----GPPSGKTGLTIAL 623
Db 937 LTTEDTRKTANLALIEDATEVGRACGCTTSP--GIVHYQFAGVSGSKRSI----- 990
QY 624 EAAAGKALYVAPTEBLREAMDRIKPPSASATOHVA--LAILRATAGAPFATVVIDEC 681
Db 991 -QGDVDVAVVPTRELNSMR--GFAAFTPHTAARVTIGRR-----VVIDEA 1036
QY 682 FMFPLVYVAVHAISSPSRIYLVGDVHOIGFIDQGSANMPLVADVVKCRRRTFNQTK 741
Db 1037 PSLP-PHILLH-MORASSVHLGDPNOIPAIDEFHAGL-VPAIRP--ELAPTSWXYTH 1091
QY 742 RCPADVATTFPGSLYPCCTTSGCVASISHVAPYRNSOQTCFTQEBKSRHGAEGAM 801
Db 1092 RCPADV--CELIRGAYPKIOTTSRYLNSLFMNER--AIGOKLYXTAALAA--PGAI 1143
QY 802 TVHEAGRTFASVILHYNGSTABOQLAEKS--HLVGIETHNMLYIRDEPTGIDBROLN 859
Db 1144 TVHEAGRTFETTTI--ATADARGLQSSRAHVALTHTKCVILDAFG--LLREVG 1199
QY 860 HS-----AKAEVFTDIPAPLEITTVKPSSEYQNR--EWMAITPPQSATPHGAIHLR 909
Db 1200 ISDVIYNNFFLAGGEVGHXP-----SVIRPGNDQMLGTLQAFPPSCQI--SAYHOLA 1251
QY 910 KNFGDQPPCGVALAKTGEVFGGRAKINVELAEADAPKPHRANQ--BGVQWVYNTASN 968
Db 1252 BELGRP-APVAAVLPCEBELQGLLYMPQELTVSDV---LVFELTDIVHCRPAADSQ 1306
QY 969 KQOALQTLISRYTKSADLPHE--AKEDVKEMLSLDHMDWTVEDARAVEFTQLKF 1027
Db 1307 RKAIVSTIVGRYGRRT---KLYEAHSDVR---BSLARFIFTIGVQATTCGLYELVEAM 1360
QY 1028 TORGGTVEDLLEPD--DPYIRDIDFLMKTQOK--VSPKPIYNGKVGOGIAHSKSLNFTLA 1084
Db 1361 VEKQDGSVAVLELDICNRDVSRIITFPQCKCNKFTTGERTIAGKVGQGISANSKTFCALFG 1420
QY 1085 AMIRILE--EILRTGSRTRYNGSLRDEBEAMLEKINOVHATVSDMTEBPDAHNT 1143
Db 1421 PMFPAIEKEITALLPENNIFYG---DAYESVFAAIVSAGSCWVFENDFSEPDSTQNNP 1476
QY 1144 SELFAALLERIGTPAAAVNLFRERCGRITLRA-----KGLGSVEVDGLDLSGAAMTPCR 1198
Db 1477 SLGLECVMEBCGMQMLIRLYHLVRSAMILLQAPKESLKGFKKH-----SGEGCTILM 1530
QY 1199 NTIFSAAVMLTL--FRGVKPAAFKDDG--LIGSHYLPDASRLHMERRYTKHLKEVQ 1255
Db 1531 NTVMMAIIAHCEFRDPRVAFAFKGDDSVLCSDRQSRNMAALIAAGCLK---LKVYDR 1587
QY 1256 KIVPYIGLVGAQVVL--DPVRSALKIFGRCTSESLYSKYVEARBDITTKGSDARYHSL 1314
Db 1588 PIGLVAGVAVPAGIGTLDPDVAFRA---GR--LSB-----KMMGP----- 1621
QY 1315 LCHMSACYNYAPBSAAVYIADVAVFCRG 1343
Db 1622 -----GPEBAQRLALVCPFLRG 1639

```

```

RESULT 6
AA131384
ID AA131384 standard; Protein; 1708 AA.
XX
AC AA131384;
XX
DT 12-OCT-1999 (first entry)
XX
DE HBV-US2 ORF1 protein.
XX
KM Hepatitis B virus; HBV; binding partner; virus; US-HBV infection;
XX vaccine; passive immunisation.
XX
OS Hepatitis B virus.
XX
FH Key location/Qualifiers
FT Misc-difference 322
FT /label= unknown
FT /note= "encoded by CRG"
FT Misc-difference 331
FT /label= unknown
FT /note= "encoded by RAC"
FT Misc-difference 445
FT /label= unknown
FT /note= "encoded by RAC"
FT Misc-difference 448
FT /label= unknown
FT /note= "encoded by RTG"
FT Misc-difference 634
FT /label= unknown
FT /note= "encoded by CCY"
FT Misc-difference 646
FT /label= unknown
FT /note= "encoded by GCS"
FT Misc-difference 811
FT /label= unknown
FT /note= "encoded by TKT"
FT Misc-difference 1531
FT /label= unknown
FT /note= "encoded by TAY"
FT Misc-difference 1578
FT /label= unknown
FT /note= "encoded by RGC"
FT Misc-difference 1691
FT /label= unknown
FT /note= "encoded by GAR"
XX
PN W09919732-A1.
XX
PD 22-APR-1999.
XX
PF 15-OCT-1998; 98WO-US21941.
XX
PR 15-OCT-1997; 97US-0061199.
XX
PA (ABBO ) ABBOTT LAB.
XX
PI Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GJ;
DR WPI, 1999-288017/24.
XX
N-PSDB; AA200267.
XX
PS Claim 6; Page 216-221; 260pp; English.
XX
The invention provides a method for detecting a US (sub)type hepatitis B
virus (US-HBV), or its naturally occurring variants in a sample by
treatment with a binding partner specific for a marker of the virus, and
then detecting any complex formed. The method is used to diagnose
infection with US-HBV. Polypeptides from US-HBV, antibodies specific for
open reading frames (ORF) in US-HBV and host cells expressing these ORFs

```

CC are useful in vaccines or for passive immunisation. The polypeptides are
 CC also used to raise specific antibodies (useful as immunoassay reagents).
 CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
 CC usual hybridisation and amplification assays for detecting infection. The
 CC present sequence represents a HEV-US2 ORF1 protein.

XX Sequence 1708 AA;

Query Match 5.8%; Score 512.5; DB 20; Length 1708;
 Best Local Similarity 21.6%; Pred. No. 4.5e-30;
 Matches 393; Conservative 186; Mismatches 557; Indels 687; Gaps 83;

13 AADVVAVVAVLQORAVVLDFAAPLKALET--LHRLVYP--LRFKGGTLPPTQPILAGHQ 68
 22 AANSLANAVVVR-----PFLSRVQTILINLMQPROLVRRPEVL--WNPI-----Q 68
 QY RVAEVLVHNFARGRS--TVLEIGPSLSHSLKLGAPNAPVADYHGC--TKYGRDGSRHITA 126
 69 RVIHNELEQYCRARAGCLEVGAHPRS---INDNPV---LHRCFLRPVGRDVGQWYSA 121
 DB 69 RVIHNELEQYCRARAGCLEVGAHPRS---INDNPV---LHRCFLRPVGRDVGQWYSA 121
 QY 127 LE-----SRVATGRPREPKADASLLANGIASRTFVDVGCAPKSRVGIANHSLYDV 179
 122 PTRGPANCRRSALRGLP-----VDRYCGDFGSRCAFAETGVALYSLHDL 169
 DB 180 TLEELANAFENHGLHWRAPFMHMPBELLVMDNVNAELGYRFVIEEPMVAVKCAFGCD 239
 QY 170 WRPDVAEAMARHGMTRLYALHLPREVLPRGT-----YHTST----- 208
 DB 240 LRLHPELDFINESQERRIERLARSGYSRRAVIFSGDDMDGAYLHDFHTMLAYL--- 295
 QY 209 LLIH-----DGNRAVVTYEGDTSAG--YHNDVSLRAMIRTK 244
 DB 296 LVANNYPTPFPSLHIEVQRHSSIELRTRAP-----PGDRMLAVVPRTSGGLCRI 347
 245 IVDHP-----LVIERVRAIGCHFVLLTAPEPSMPYVPYPRSTEVYRSIFPGGS 298
 QY 348 PNIFFYADAGTEHKTILTQHKVNMMLNMQTRPEKELVDMVWMSFAPARALAIYVAS 407
 DB 299 PSLEF--PSACSTSTFIAVAVHWDKMLMFGATLXDAPFC--CSRLMTYLRGISYKYVGA 355
 QY 408 EYTESWNISPADLVRTVSLYYLHIIERRAAVAV----- 443
 DB 356 LVANEGWNASIEDALTAVITAYLITICORYLRTQALSKMRLEVEHAQKITRLYSWLF 415
 QY 444 -KRAKDV-----FGETSFWS-----LKVYLS--CCGLR 471
 DB 416 EKSGRDYIPGRLOFYAQCRKMLSAGFLXPRXLVPDESVPCCRRTFLKXVAGFCFMR 475
 QY 472 NL-----KGTDV----- 478
 DB 476 WLGQECTFLPBAGLVGDQGHNDNEAYEGSEVDPAEPAHLDVSGTVAHGHOLEALYAL 535
 QY 479 -----VETKEVVDKRYV----- 490
 DB 536 NVPHDIAARASRLTATVELVASPDRLRCRTVLGNKTRRTTVODAHLEANGPBEYVLSFD 595
 QY 491 -----HSL-----GDI--ICD----- 499
 DB 596 ASROSGMGASHSLTYELTPAGLQYKISSNGLDCTATEPGGAPSAAGBYAXAFCSALYRY 655
 QY 500 -----VRLSPQOV-----GFLPSRV----- 514
 DB 656 NRFTQHSLSLGLMLHPEGLIGTFPPSPGHIMESANPFCGEGTLYTRTWSGFSDFS 715
 QY 515 -----PPARVFN--DREBLEVLARE--AGCYNE--RPV 540
 DB 716 PPEAAAPASAAAPGLPYPTPVPVSDIWLPRPSESHTDASVSPVBPAGLTSPIVLTPP 775
 QY 541 PSTPPVEP----- 549
 DB 776 PPEPPVKKPATSPPTTRRLLYTYPDGAKYVAGSLKESDCMLVNASNPGHRPGGGLCHA 835
 QY 550 -----QGFDV-----DLWHTATPAASLPEYRA-----TLQAG----- 574

DB 836 FYGRFPEARVSTEFIMREGIAVYTLPRRIHAVA--PDYVNBQPKRLBAAYNETCSR 892
 QY 575 -----LNTVKKQKLT----- 588
 DB 893 RGTAAVPLIGSGIYQVPVSLSPDAMERNRPPDBYLTERPAAMPEANKPAPALTTTID 952
 QY 589 ALKT-----ID-----GLTSPVRLGEMTE--GPPSGGTGTLIAALEAAGK 629
 DB 953 TARTANLALIEIATAEVBGACGCTISP--GIWHYQPTAGVDSGSKRSI-----QQGV 1005
 QY 630 ALVVAETRELRAMDRIRIPESATQHYA--LALIRATAGAPATVVIDECMFPLV 687
 DB 1006 DVVVAVPTRELSMRMR--GPAATPHRAAVTIGRR-----VIDEASLP--P 1051
 QY 688 YVALYHALSPSSRIYLVGDVHOGITDFGCTSANPPLVDVVKQCRRTFNQTKCPADV 747
 DB 1052 HLLLLH--MORASSVHLGDPPNOIPADPEHAGL--VPAIRP--ELAPTSMMWYTHRCPADV 1107
 QY 748 VATTFPQSLYPGCTTSGCVASISHVAPYRNSQAOTLCFTQEBKSRHGAEGMTVHEAO 807
 DB 1108 --CELIRGAYPKIQTTSRVLRLFMNRP-----AIGQKLVTOAKAAN--PGALIVHEAO 1159
 QY 808 GRTPASVILHNGSTABQCLAEKS--HLLVGITRTHNLVTRDPTGDIROLNHS----- 861
 DB 1160 GATFTETIT--ATDARAGLIQSRBAHVALTHTKCEVILLAPG--LIREVGISDVIV 1215
 QY 862 -----AKAVFTDIPAPLEITTVKPSSEVQRN--EVMATTPROGATPHGAILHLKRNFGQ 915
 DB 1216 NNFFLAGVGEVGNRP-----SVIPGNPDQMLGTLOAPFPCOI--SAYHOLAEBELGHR 1267
 QY 916 PDGCVALLAKTGVYEGGRAKINVELABDAPKPHARFO--BGVQVVKVTNNSNKHQALQ 974
 DB 1268 P-APVAVALPSCBELGGLIYMPOLITVSDV--LVLELDIVHCRMAABSORAVUS 1322
 QY 975 TLLSRYTKRSADLPLHE--AKEDVKRMANSLDRHMVTTVEDARDRAVFTOLKFTORGST 1033
 DB 1323 TLVGRYGRRT--KIYEABHSVDR--ESLAPIPTIGVRAATTCELYELVEAMVEKGDD 1376
 QY 1034 VEDLEPDD--DPIYRIDIDPLMTQOK--VSPKPIYNGKVGQGIAMSKSLNPFVLAIRLL 1090
 DB 1377 GSAVLELDICNDRDVSRTTFPODCNKFTTGERTIAHGKVGQGISMSKTCALFGPWFRAI 1436
 QY 1091 E-BILFTSGRTVRSNGLPRDEBAMLEAKINQVPHATVSDMTFEDFAHNNTSELPA 1149
 DB 1437 EKSLIALLRPNIIFYG---DAYBSVFAALVSGAGSCMFPENDFSFSDTONNFSGLHC 1492
 QY 1150 ALLERIGTPAAAVNLFRERCGRKTLPA-----KGLGSVEVDGILDSGAAMTFCRNTIFSA 1204
 DB 1493 VVMEBCGMQMLRLVLRSAWILQAPKBSLKGFWKCH-----SGERGTLIMNTVNM 1546
 QY 1205 AVN--LTLFRGVFAAPKDDDS--LICSGHYLRFDSARLMGEBRYTKHLKVEYQKIVPI 1261
 DB 1547 AIIAHCKBPRDFVAAFKGDDSVLCSDRQKRNMAALLIAGGLK--LKVDYRPIGLYA 1603
 QY 1262 GLLVSAEQVVL--DPVUSALKIGRCTSELILSKYEAARVDITKGSDDARHSILCHMA 1320
 DB 1604 GVVAAPGLGTLPLPVAFPA--GR--LSR-----KNMP----- 1631
 QY 1321 CYTVAPESAAAYIIDAVNFRG 1343
 DB 1632 -----GPERABQRLAIVCDPLRG 1649

RESULT 7
 AAM93408
 ID AAM93408 standard; Protein; 1708 AA.
 XX
 AC AAM93408;
 XX
 DT 11-JUN-1999 (first entry)
 XX
 DE Swine HEV ORF 1 protein.

QY 1028 TORGTVEVDLLEPD--DPYTRIDIDFLMKTQOK-VSPKPIINTGVGCGIAAHSKSLNFVLA 1084
 Db 1371 VEGGQDGSAYLELDLCNRDVSRTTFQKCNKKTGTGETIAHGIVGCGISAMSTFGALPG 1430
 QY 1085 AMIRIDE-ELTRGSTRVRSNGLPDEEAMLEAKINQVPHATFVSADTEPTAHNNT 1143
 Db 1431 PWFRAIEKEILLAPRNIIFG---DAYESVPAALVSGAGSCMVENDSEPTDSTQNNF 1486
 QY 1144 SELFAALLERIGTPAAAVNLFRRCQKTLRA-----KQIGSVEVDGLDLSGAATPGR 1198
 Db 1487 SGLGECVMBECMPQMLIRLHYLVRSAMILQAPKSLKGFMKH-----SGEPGLLM 1540
 QY 1199 NTFSAAVMLTL--FGVKPAAPKGDSD-LLCGSHYLRPDASRLHNGERYKTKHLYEVO 1255
 Db 1541 NTVMNNAIIAHCEYFRDFRVAAPKGDSDVVLCSDYRQSRMAALIGCGLK--LVVDYR 1597
 QY 1256 KIVPYIGLVSARQVVL-DPVRSALKIFGRCTSELLYSKYVAVRDITKMSDARYHSL 1314
 Db 1598 PIGLIVGVVAVPGLGLTLPDVVRA---GR--LSE-----KNNGP----- 1631
 QY 1315 LCHMSACYNVYAPESAAVYIIDAVVRGKG 1343
 Db 1632 -----GPERAEQLRLAVCDFLRG 1649

RESULT 8
 AAW76368
 ID AAW76368 standard; Protein; 1693 AA.
 XX
 AC AAW76368;
 DT 03-DEC-1998 (first entry)
 XX
 DE Hepatitis E virus hollow particle protein #1.
 XX
 KM Hollow particle protein; virus; antibody; detection; immunoassay;
 XX
 OS Hepatitis virus.
 XX
 FH Key Location/Qualifiers
 FT Protein 1..1693
 /note="Partial sequence"
 XX
 PN JP10234383-A.
 XX
 PD 08-SEP-1998.
 XX
 PF 28-FEB-1997; 97JP-0062445.
 XX
 PR 28-FEB-1997; 97JP-0062445.
 XX
 PA (DENK-) DENKA SEIKEN KK.
 (KOKU-) KOKURITSU YODO EISEI KENKYUSHO.
 DR N-PSDB; AAV61687.
 XX
 PT Hepatitis E virus hollow particle poly-peptide(s) and nucleic acids
 encoding it - useful for more accurate detection of HEV in samples,
 using immuno-assays and nucleic acid hybridisation
 XX
 PS Claim 10; Page 17-24; 29pp; Japanese.
 CC This sequence represents a Hepatitis E viral hollow particle protein.
 CC This polypeptides can be used to raise antibodies to detect HEV
 CC infection in samples, e.g. by immuno-assay based techniques, and the
 CC nucleic acid can be used for the same in nucleic acid hybridisation
 CC assays. The polypeptides and nucleic acids allow more accurate
 CC detection of HEV than previously possible.
 XX
 SQ Sequence 1693 AA;

Query Match 5.7%; Score 506; DB 19; Length 1693;
 Best Local Similarity 21.4%; Pred. No. 1.4e-29;
 Matches 375; Conservative 190; Mismatches 559; Indels 630; Gaps 75;

QY 2 YKATDVAAVVAADVAVVNVVQRAVYKDFAPPLKALTELRLYVP--LRKGGTLPT 59
 Db 6 FTKAPGTTATIGQALAAANSLAANAANVVRPLSHQOITILINLQPROLVFRPEVY--W 63
 QY 60 QHPILAGORVABEVVHNFARGRS--TVLEIGPSLSALNLAGAPNAPVADYHGC--TKYGT 117
 Db 64 NHPI---QRVHNEBELVCRAISGRGLEIGHAPS--INNPNV---VHRCFLRPVG 112
 QY 118 RQGSRHITALE-----SRVATGRPEPKADASLNGIASRTFCVDGSCAPKSRVG 170
 Db 113 RVQWRVYTPAPTGPANCRSRALRGLP-----AVDRYCLDGPSCGNFPAEYG 160
 QY 171 IANHSLYDVLTEELNAPENHGLHNVRAFMHPEBLLYMDVNVNMLEGYFVIEBPAV 230
 Db 161 IALYSLHDSPSDVAEMFRHQHTRLYAALHLPPEVLL-----PGRTY 203
 QY 231 KDCAPQGDRLHFPELDPIFESQERRIERLARAGSYSHRAVIFSGDDDMGDAYLHD--- 287
 Db 204 RTASY-----LLH-----DGRV-----VVTYEGDTSAG--YHNDVSN 235
 QY 288 FHTWLAIVLVKNYPTPPGSLHIEVORRHGSSIELRTPAP-----PGRMILAVPR 339
 Db 236 LRSM-----IRTKVYGDHPLVIERVRAIGCHFVLLTAAPEPSMPYVPRSTEVYR 290
 QY 340 TSGGLCRIPNIFYVADASGTBHKTLTSGHKNNMLNFMQTRPEKELVDMTULMSFAR 399
 Db 291 STFGPGTSLP--PTGCKSTFHAIVPAHWDRLMFGATLDQAF--CSRIMTYLRGI 347
 QY 400 LPAIVASVETESSWNISPADLVRTVSLV-----LHIERRAAVAV 443
 Db 348 SYKVTYGLIVANGMNASDALTAIVITAYLITCHQRYLRTQAISKGRRLERHOKXI 407
 QY 444 -----KTADVFQ--ETSPWESLKIVLS----- 466
 Db 408 TPLYSWLFEKSGRDYIPGRQLFEYACRRMLASGFLDPRVLVPDESACRCRTAIRKAL 467
 QY 467 ---CC-----GLRLKGTDVYPTGR 483
 Db 468 SKFCFPMKMLQGECTCFQAPAGVGDQCHNEAVEGSDVPAESAIISDSYVVGTA 527
 QY 484 VVDKYRVHSL-----GDILCDVRL-----SP 504
 Db 528 LQPLVQALDLPABIAVARGLTATVVGQVDRIDCETILGKKTPTSFDGAVLEANGP 587
 QY 505 EOVG-----FLPS---RVPPARV-- 519
 Db 588 ERYNLSFDASGTMAGPSPTYAASAGLEVRYVAAGLDHRVAFAPGVSPSAPGEVTA 647
 QY 520 -----FHD----- 522
 Db 648 FGSALYFRREARQHSLSGNLMFHPGLIGLAPSPGAVMESANPCGSESTLYRTWSB 707
 QY 523 -----REBLEVLEACYNR-----PVSTPYE-- 548
 Db 708 VDAVSPAPRDLGLMSEPSIIPSRATPTLAPLPLAPRPPPSAPALDEPAGSATAGA 767
 QY 549 -----PQ 550
 Db 768 PAITQTAHRRRLFTYPPDSKVPAGSLSESACTWLVNANVDHRRGGGLCHA FYORYP 827
 QY 551 GPDA-----DLMHATAASLPEYR-----ATLQAG----- 574
 Db 828 SFDASFPWRDGAAYTLTPRPIHAVA---PQYRLEHNPCKLEAAVRETSCLGTAAYP 884
 QY 575 -LNTDVQK-----L 582
 Db 885 LIGTGIVYVPIGSPFAMERNHRPGDELYLPBLAARWFANRTRPTLTITTEDVARTANL 944


```

Qy 484 VDKYVHSP-----GDICDVR-----502
Db 528 LQPLVALDLPAIIVARAGRLTATVKSQVDGRIIDETLIGNKTRTSFVDAVLSTNGP 567
Qy 503 -----502
Db 588 ERHNLSPDASQSTMAAGPFLTYAASAGLEVRYVAAGLDRAVAPGVSPGAPGVTA 647
Qy 503 -----502
Db 648 FCSALYRFRREARQLSLTGNFMWHPBGLGPAPFPGHWESANPFCGBSTLYRTWSR 707
Qy 503 -----502
Db 708 VDAVPSAPQDLGFTSEPSIPLSAATPTAPAPLPPAPDPSPPLSAPARGBAPGATARA 767
Qy 518 -----517
Db 768 PAITHTARHRLLFYPDGSKVFAGSLFESTCTWLNVASNDVRPGGGLCHAFYQRYPA 827
Qy 518 -----517
Db 828 SPDAFVWRDGAAYTLTPRPPIHVAAPDYRLHNPKRLAAVRETCSRLGTAAVPLIG 887
Qy 543 TPVEBPQGFADLMH-----ATAASLPEYRA-----TLQAGLTDVYKO---LK 583
Db 888 TGIYQVPIGPSFAMERNRPGDELYLPELAAMFEANRPTCTLT--ITEDVARTANLA 945
Qy 584 ITLENAL---KTIDGLTSPVRLGEMYE---GPPGSGKTGTLTIALEAAGKALVYAPTR 637
Db 946 IELDSATDVGRACAGCRVTP--GVVOYQFTAGVPGSGKSRSTQA-----DVDVVVPR 998
Qy 638 ELREANDRIKPPSASATQVALAIL--RRATABGAPATVVIDDECPMLPVVAIVHAL 695
Db 999 ELRNAMRRR---GFAAFPTPTATARYTQGR-----VVIDEAPSLP-PHLLHLH-M 1043
Qy 696 SPSSRIVLVDVHIOGIFIDFGTSAMPLVADVVKCRRTFQTRCPADVATTEFOS 755
Db 1044 QRAATVHLGDPRQIPIAIDFENAGL--VPARIPDL--APTSWMTVTRCPADV--CEILING 1098
Qy 756 LYPGCTTSGCVASISHVAPDYNSQAQTLCTFOEBSKSRHAGAMTVEHAGRTFASVY 815
Db 1099 AYPMTIGTGRVLRSLFMGEB---AVGQKLVFTQAANKAN--PGSTVYHAGQATYETT 1152
Qy 816 LHYNGSTAEOKLAES--HLVGITRHTNHLTYIRPTGDIERQLNHSKAEVFTDIPAP 873
Db 1153 II---ATADARGLIQSSRAHAIVALTTRHEKCVIIDAPG-LIREVGIS--DAIVNPFPLAG 1207
Qy 874 LEITTVKPSBEVORN-----EVMATIPPOSATPHGAILLRKNFGDOPDGCVALAKTGV 928
Db 1208 GEIGHORPSVYIPRGNDANVDITLAAPPSCEI--SAFHLEAELGHRP--APVAVALPPCP 1264
Qy 929 EVFGGRAKINVELAEDATPKPHRAFO--EGVQWVKVTNANKNHQAQTLISRYTKSADL 987
Db 1265 ELEQGLLYVPOELTTCDVV---TFELTDIVHCRMAAOSQRAVSLTVLGRGRT--- 1317
Qy 988 PLHEAK--EYVKMLNLDHRMDWTVEDARDRAVEFTOLKFTORGCTVEDLEPD--DRY 1044
Db 1318 KLYNASHSDVR---DLARFIPAIGVQVTTCTELYELEBEAMVEKGDGSGVELDSCSD 1374
Qy 1045 IRDIDIMLTKTOOK--VSPKPIINGKVGQGIAAHSKSNFLAAMIRLLE--ILRTGSRTRY 1102
Db 1375 VSRITTFQDKCNKFTTGETIAHCKVQGISAMSKITCALPGPFRAIEKAILLAPQGVF 1434
Qy 1103 YSNGLPDEE--EAMLEAKINOVPHATFVSDWTEPDTAHHNTSELFAALLERIGTPAA 1161
Db 1435 YGDAPFDVFSAAVAAKASMV-----FENDFSEFSTQNFSLGLECAIMECGAPQWL 1489
Qy 1162 VNLFRRCGRKTLIRAGLGSVEVDG---LDSGAAMPCCRTTISAAVMLTL--FRGV 1214
Db 1490 IRY-----HLIRSAMILLQAPKESLRGFWKXSGBPGLTLMNTVMNMVITHCYDFDL 1543
Qy 1215 KFAAFKGDSSL--LCGSHYLRFDASRLHMGERYTKHLKVEVQIVYIGILVSAEGQVL- 1272

```

```

Db 1544 QVAAPFKGDDSYLCSGTSRSPAAVLIAGCGK--LKVDPRIGIYAGVVAAPGALP 1600
Qy 1273 DPVRS 1278
Db 1601 DVVRF 1606

RESULT 10
ID AAM81519
AA81519 standard; Protein; 1693 AA.
XX
AC AAM81519;
XX
DT 02-FEB-1999 (first entry)
XX
DE Hepatitis B virus (HEV) ORF-1 protein.
XX
KM Hepatitis B virus; HEV; SAR-55; diagnostic agent; vaccine; antibody;
XX passive immunisation; open reading frame; ORF.
XX
OS Hepatitis B virus.
XX
FH Key Location/Qualifiers
FT MISC-difference 1238
FT MISC-difference /note= "encoded by CAG"
FT MISC-difference 1244
FT MISC-difference /note= "encoded by CAG"
FT MISC-difference 1352
FT MISC-difference /note= "encoded by GTC"
FT MISC-difference 1662
FT MISC-difference /note= "encoded by GGC"
XX
PN WO9846761-A1.
XX
PD 22-OCT-1998.
XX
PF 09-APR-1998; 98WO-US07418.
XX
PR 11-APR-1997; 97US-0840316.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX
PI Emerson SU, Purcell RH, Robinson RA, Tsarev SA;
XX
DR WPI; 1998-568733/48.
XX
N-PSDB; AAV71604.
XX
PT New hepatitis B virus DNA from Pakistani strain SAR-55 - used for,
PT e.g. developing products for diagnosis of, and vaccination against
PT hepatitis B virus infection
XX
PS Disclosure; Pages 11-15; 204pp; English.
XX
XX
This represents a hepatitis B virus (HEV) open reading frame (ORF)-1
protein encoded by a DNA sequence designated SAR-55. SAR-55 also encodes
HEV ORF-2 and ORF-3 proteins. A host organism transformed or transfected
with a recombinant expression vector containing the SAR-55 nucleic acid
can be used to produce the HEV proteins, especially ORF-2 protein. The
recombinant HEV proteins can be used as diagnostic agents and as vaccines
for use against HEV infection. The detection of antibodies specific for
HEV can be used for the diagnosis of infection and diseases caused by
HEV, and for monitoring the progression of such disease. Such methods are
also useful for monitoring the efficacy of therapeutic agents during the
course of treatment of HEV infection and disease in a mammal. The
antibodies can be used for detection or for passive immunisation of
mammals.
XX
SQ Sequence 1693 AA;
XX
Query Match 5.7%; Score 500; DB 19; Length 1693;
Best Local Similarity 21.2%; Pred. No. 4.3e-29;
Matches 371; Conservative 186; Mismatches 575; Indels 614; Gaps 72;

```



```

QY      2 YAKATDVARYAADAADVAVANVLQORAYKLDPAFLKALETHRLTYR--LRFKSGTLPPT 59
      6 FIKAPGTTTALIEQALALAANSALANAAVAVPRLSHQOIEILINIMOPQOLVREVF--W 63
QY      60 QHPLIAGHORVAEEVLNHFAGRS-TVLEIGPSLSALKHGANAPVADYHG--TKYGT 117
      64 NHPL-----QVINEHELYCAASGRCLIEGAPRS--INDNRV-----VHRCFLRAG 112
QY      118 RDGSRHITALE-----SRSVATGRPEFKADASLANGLASRTFCVDGVSCAFKSRVG 170
      113 RDVGRWYTAFTPRGPAACRSALRGLP-----AADRTYCFDGSFGCFPRLETG 160
QY      171 IANHSYDVTIEELANAFENHGLMVRAPFMPEBELLMYNDVNAELGYRPHVIEPMAY 230
      161 IALSLHDMSPSDVAEAMFRHGMTRLYAALHPREVL-----PGTY 203
QY      231 KDCAFQGGDLRLHPELDFINESQBRRIERLAAGSYSRRAVIFSGDDMDAYLHD-- 287
      204 RTASY-----LIIH-----DGRV-----VITYEGDTSAG--YNHDSN 235
QY      288 FHTWLAVLVNRYPTPGFSLHIEVORHSSIELRTTRAP-----PGDRMLAVVPR 339
      236 LRSM-----IRTKVTGDHPLVIERVRAIGCHFVLLLTJABEPSMPYVPRSTGYVR 290
QY      340 TSQGLCIPIFYIADASGTEHKTILTSQHKVMNLNFMOTRPEKELVDMTVLMSFAPAR 399
      291 SIFPGGTPLF--PTSCSTKSTFHAIPAHIWDRMLFGATLDOAFC--CSRLMTYRGI 347
QY      400 LRAIVASEVTESWNISPADLVRTVSVLVY-----LHIERBRAAVAV 443
      348 SYKTVGTGLVANEEMNASSEDALTAVITRAVYLTICHQRLTQALSKMRRLEREHAQKI 407
QY      444 -----KTAKD-----DVEGETSFWESE-----KHV 463
      408 TRLSWLFKSGRYIPGRQLEFYAQCRMLSAGHLDPRVLVDESAPCHCTAIRKAV 467
QY      464 LGSCC-----GLRNKGTDVVFTKR 483
      468 SKFCFPMKWLQOECTCLQPAEAGVVGQGHNEAVEGSDVDPAESAISDISGYVVGTA 527
QY      484 VVDCYRVHSL-----GDIIQVRL----- 502
      528 LQPLIYQALDLPALIVARAGLTAIVKVSQVDGRIDCETLGNKTFRTSFDGAVLETNGP 587
QY      503 ----- 502
      588 ERHNLSDASQSTMAAGPFSLTYYAASAAGLEVRYVYAGLDRHAFAPGVSPRSAGEVTA 647
QY      503 ----- 502
      648 FCSALYFNRBAQRLSTGNFMFPEBGLGPPAPFSPGHVWESANPFCGESTLYTRTWE 707
QY      503 -----SPEQ--VGF-----LPSRV-----PPA----- 517
      708 VDAVPSAQPDPDGTSEPSIPSRATPTPAAPLPPAPDPSPTLSAPARGPAPGATARA 767
QY      518 ----- 517
      768 PAITHQTAHRRLFTYPDGSKVFAAGSLFESTCTMLVNASVNDHPPGGGLCHAFYQRYRA 827
QY      518 -----RVPHDEELVLABAGYN-----ERPVS 542
      828 SFDASFVMDGAAYTLTPRPIIHAVAPDYRLBNPRLEAAVRETSRLGTAAYPILG 887
QY      543 TPVYEEPGFADLWM-----ATAASLPEYRA-----TLQAGLNTDVQO---LK 583
      888 TGIYQVPLIGSPFDAMERNHRPBGDELYBELAARPEANRPICPLT--ITBDVARTANLA 945
QY      584 ITLENAL---KTIDGLTSPVRLGEMYE---GPPSGKGTGLTIALAELAAAGKALYVAPTR 637
      946 IELDSATDVGRACAGCRVTP--GVVQYQFTAGVFGSGKRSITQA-----DVDVVVVPTR 998

```

```

QY      638 ELREAMDRIKPPSASATQVALIL--RRATAGAPFATVVIDECMPFLVVAIVHAL 695
      999 ELRNAMERR--GRAATPHTAARVTQGR-----VIDEAPSLP-PHILLH-M 1043
QY      696 SPSSRIVLVDVHQIGFIDPQGSANMPFLVVDVVKQCRRTPTQTKCPADVAATPFGS 755
      1044 QRATVHLLDDPQIPRIDEBHAGL--VPAIRPLP--APTSWVHTHRCPADV--CEIIRG 1098
QY      756 LYPCCTTSGCCVASISHVADYRNSAQTLCTFOEKSRRHAGAMTVEAOGRTFASVI 815
      1099 AYPMIQTTSVHLSLFGEBP-----AVGQKLVFTQAALAAAN--PGSVTHBAQGATVRETT 1152
QY      816 LHNQSTAEQKLAELS--HLVGIYRTHNLVIRPTEGIERQNLHSAQAEFTDIPAR 873
      1153 IT---ATAADRGLIQSRAIAIVALTRHEKCVIAPG--LIREVGIS--DAIVNPFILAG 1207
QY      874 LEITTVPSSEEVORN-----EVAATIPQSATPHGAHILKRKFGDQPCGCVALAKTG 928
      1208 GEIGHQRPVTPRGNPANDDTLAAPPSCET--SAFHELABELGHRP-APVAAVLPPCP 1264
QY      929 EYFGRAKINVELAEPDAPKPHRAPO-EGVQWVKVTNASNKQIQLTLLSRYTKRSADL 987
      1265 ELEQGLLYLPQELTTCDSV-----TFELTDIVHCRMAAPQRAKAVLSTLVGRYGRAT--- 1317
QY      988 PLHEAK-EDYKRLNLSDRHMDWTVEDADRAVFEQLKFTQRGCTVEDLLEPD--DPY 1044
      1318 KLYNASHDVR---DSLARTIPALGPVQVTTCELYELREAMEVEKQDGSANLELDLSRD 1374
QY      1045 IRDIDPLMKTOOK-VSPKPINTGKVGQGLAHSKSLFVLAAMRILIER--ILRTGSRTRV 1102
      1375 VSRITFGQXONKFTTGETIAHGKVGQIGIAMSKEITPCALGMPFRAIKEXIALLLPGCVF 1434
QY      1103 YSNGLPDEE-BAMLEAKINQVPHATVFSADWTEFDTAANNSTELLEALLERIGTPAA 1161
      1435 YGDAFDDTVPSAAVAAKASMV-----FENDPSEBFDSTQNNPFSIGLECAIMEBCGMOWL 1489
QY      1162 VNLFRERCGKRTLRAKLGSVEVDGL-----LDSGAAMTCRTTIPSAAMVLT--FRGV 1214
      1490 IRLY-----HLIRSAMITQAPKESLNGFMRKKSISGEGTLLMTVMMAVITHCYDRDL 1543
QY      1215 KPAAFKGDSDL-LGSHYLRFDASRLMGRYTKHLKKEVQKIVPIYGLVLSAEQVVL- 1272
      1544 QVAAFKGDDSIIVLCSEYRQSPGAANVLAGGLK---LKVDPRIGLYAGVAVPAGLALP 1600
QY      1273 DPVRSA 1278
      1601 DVVRFA 1606

```

RESULT 11
 AAR91813
 ID AAR91813 standard; Protein; 1693 AA.
 XX
 AC AAR91813;
 XX
 DT 26-NOV-1996 (first entry).
 XX
 DE Hepatitis E virus strain SAR-55 ORF-1.
 XX
 KW Hepatitis E virus; HEV; SAR-55 strain; enteric transmission;
 KW structural region; antigen; detection; antibody; vaccine;
 KW immunization; infection.
 XX
 OS Hepatitis E virus.
 XX
 FH Key Location/Qualifiers
 FT MISC-difference 1238
 FT MISC-difference /note= "corresponding codon CAG"
 FT MISC-difference 1455..1693
 FT MISC-difference /note= "10 bp nucleic acid sequence TGCATTGGA
 has to be inserted between nucleotides
 4390..4391 of AAT72394 before these amino
 acid residues can be decoded"
 FT

```

XX  MO9610580-A2.
PN
XX  11-APR-1996.
PD
XX  03-OCT-1995; 95WO-US13102.
PF
XX  03-OCT-1994; 94US-0316765.
PR
XX  (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA
XX  Emerson SU, Purcell RH, Tsarev SA;
PI
XX  WPI: 1996-209320/21.
DR  N-PSDB; AAT27394.
XX
XX  Isolated and purified hepatitis E virus strain SAR-55 DNA - encodes
PT  antigenic protein useful in diagnosis, prophylaxis and treatment of
PT  hepatitis E virus infection
XX
XX  Disclosure; Pages 9-13; 121pp; English.
PS
XX  The present sequence is the protein prod. of ORF-1 from the
CC  hepatitis E virus (HEV) strain SAR-55, which was implicated in an
CC  enterically transmitted non-A, non-B hepatitis in Pakistan. The
CC  protein encoded by the structural region of the virus (i.e. ORF-2),
CC  which is capable of forming HEV like particles, is useful for the
CC  detection of HEV antibodies (pref. IgG or IgM) in blood, plasma,
CC  sera, cerebrospinal fluid, tissue, urine or pleural fluid. The
CC  protein, and anti-HEV antibodies generated using the protein, can
CC  also be used in vaccines for immunising an animal against HEV
CC  infection. The protein is identified as a band of greater than
CC  50 kD following SDS-PAGE of cell lysates of insect cells infected
CC  with a HEV ORF-2 conig. baculovirus, i.e. the claimed recombinant
CC  expression vectors pPIC9-1779, -1780 and -1781.
XX
SQ  Sequence 1693 AA;

Query Match      5.6%; Score 497; DB 17; Length 1693;
Best Local Similarity 21.2%; Pred. No. 7.3e-29;
Matches 311; Conservative 185; Mismatches 576; Indels 614; Gaps 72;

QY  2 YKATQVARVVAADYAVANVLOQRAVKLDPAFPLKLTILHLYTP--LRFKGLTPT 59
DB  6 FTKAPQITTAIEQALAAANALANAVVVRPLSHQIETILNMQPRQVFFPEVVF--W 63
QY  60 OHPILAGHORAEEVLHNPARGRS--TLEIGSLHSAKLHGAPNAPVADYHGC--TKYGT 117
DB  64 NHPI-----QVTHNELELYCRANSGRCLEIGHAPRS--INDPNV-----VHRCFLRPA 112
QY  118 RQGSRIITLALB-----SRVATGRPERKADASLIANGIARTFCVDDGVGSCAFKSRVG 170
DB  113 RDQQRVYTAPTRGPANCRSRALRGLP-----AADRTYCFDGGSGGNFPAETG 160
QY  171 IANSHYDTLEELANAFENHGLHWRAFHMEBELLYMDNVNNAELGFRHIEBPMAY 230
DB  161 IALYSIDHSPSPDVAAMFRHGHTRLYAALHLRPEVLL-----PQTY 203
QY  231 KQCAFQGDRLRHFPELDFINESQERRIERLARGSYSRRAVIFSGDDDDMGAYLHD--- 287
DB  204 RTASY-----LLIH-----DGRRV-----VVIYEDDTAG--YNDVSN 235
QY  288 FHTWLAVLVNRYPTPPGSLHIEVQRHGGSIETLRITRAP-----PGDRMLAVNPR 339
DB  236 LRW-----IRTKVTGDHPLVIERVRAIGCHFVLLLTAAPEBSPMPVYPFSTEVYAR 290
QY  340 TSGGLCRIPNIEFYVADASGEHKTITLSQHKVMILNFMQTREREKELVMTVMSPAR 399
DB  291 STGPGGTSLF--PFSCTKSTFFHAVPAHIMRLMLFANTLDDQFC--CSRLMTLTKGI 347
QY  400 LRAIVASEVTESSANISPADLVRTVSLV-----LHIERRRAVAV 443
DB  348 SYKVTGTLVANGMNASBDALTAIVITAVLTICHQRYLRTQALISKGMRLREBHQKFI 407

```

```

QY  444 -----KTXAD-----DVGQTSFRESL-----KHV 463
DB  408 TRLYSWLPEKSGRDYIPGRQLEFYAQCRMLWSAGEHLDPRLVLPDESAPCHRTAIRAV 467
QY  464 LGSCC-----GLANLKTQDVFTKR 483
DB  468 SKFCCEFMKLGQECTCFIQPABGVVDDQHNEAVESGDVDPABASIDISGSIVVPGTA 527
QY  484 VDKTYRVS-----GDIIQDVL----- 502
DB  528 LQPLTQALDLPABIAVARAGRLTATVKSQVGRIDCETLLGNTKFRTPSFVDAVLETGP 587
QY  503 ----- 502
DB  588 EHNLSFDSQSTMAAGPSSLTYAASAGLEVRYVAAGLDHRAPVAPVSPRABGEVTA 647
QY  503 ----- 502
DB  648 PCSALYRRENREKQRLSLTGNFMFHPEGILGPAPPSPGHVBSANPCGCESTLYRTWSE 707
QY  503 -----SPBQ--VGF-----LPSRV-----PPA----- 517
DB  708 VDAVSPAPQDLPGLFSBPSIFSRATTPPAAPLPAPADPPTLSAPARBPAGATARA 767
QY  518 ----- 517
DB  768 PAITHQTAHRRLFTYDPSGKVPAGSLFESTCTWLVAASVNDHRPGGLCHAFYQRYPA 827
QY  518 -----RVFDRBELVLRAGCYN-----BRPV 542
DB  828 SFDAASPWRDGAAYTITPRPIIHAVAPDRLBNPKLEBAVRETSLGTAAYPLIG 887
QY  543 TPVEEPOGFDADLW-----ATASLPEYRA-----TLQAGLNTDVKO--Lk 583
DB  888 TGIQVVPICPSFDAERNHRPDDELXLPELAARWFEANRPTCTILT--ITEDVARTANLA 945
QY  584 ITLENAL---KITDILTSPVRLGEMT---GPRSGTGITLIALBAAGKALVYAPTR 637
DB  946 IELDSATDVGRACACRATP--GVQYQFTAGVPSGSRSTQK-----DVDVVVTPR 998
QY  638 ELREAMDRIRKPPSATQOVNVALT--RBAATGAPATVYIDECFPLVYVAIVAL 695
DB  999 ELRNMRKR--GPAFTPHTAARTQGR-----VIDAASLP--PHLLILH--M 1043
QY  636 SPSSRIVLVGDVHQTGIDPQGSAMPLVBDVVKQCRRTFNQTKCRPADVATTFQS 755
DB  1044 QRAATVHLIDPBNQIPALIDFENAGL--VPAIRBDL--AFTSMWHVTHRCPADV--CELRIG 1098
QY  756 LYPGCTTSGCVASISHVAPDRNSQAQTLCTQSEKSRHAGAMTYHBAQRTPAVYI 815
DB  1099 AYPMTQTSSRYLRSLEWGP-----AVGQKLVETQAAKAN--PGSVYHBAQGAAYTBT 1152
QY  816 LHYNSSTABOKLLAKS--HLVGLTRHTNHLIYDPGDIERONASHAKAEVFTDIPAR 873
DB  1153 IT---ATADAKELISSRAHAIVALTRHTEKCVIIDARQ--LIREVGIS--DALVNNPFLAG 1207
QY  874 LEITTVKSEEVORN-----EYMATIPPOSATPHGAIHLRKNFDPQDQCGVALAKTGY 928
DB  1208 GEIGHQRPSVIRGNPDANVDTLAAPPSCBI--SAPQLABELGHR--APAAVALPSCP 1264
QY  929 EVFGGRAKINVEIAPBDATPKPRHAFQ--EGVQWVYVYNASNGHQLQTLISRYTRGADL 987
DB  1265 ELEQGLLYPQELTTCDSVV-----TFELDIYHCMAAPSPQKAVSLTVGVGGYRT-- 1317
QY  988 PLHEAK--EDVKMLNSLDHMDWTVEDARDAVEETQKLPORGAGTYEDLLEPP--DPY 1044
DB  1318 KLYNASHSVR---DSLARFIPAIGPVQVTTCELYELVYAMRYEKQDGSAAVELDLCSD 1374
QY  1045 IRDIDFLMKTQOK--VSPRPINTGVGGGIAHNSKSLNVLAMWIRILEE--ILRTGSRVYR 1102
DB  1375 VSRITFPQDCNKFTTGETIAHGKVGQGISASKTFCALFGWFAIERALIALALLPQGVF 1434

```

QY 1103 YSNGLPDEE-EAMLEAKINOVPHATFVSADWTEFDTAHNTSELLPALLERIGTPAA 1161
 DB 1435 YGDAFDDVFSAAVAATAASV-----FENDSESDSQNNFSLGLECAIMEBCMPQWL 1489
 QY 1162 VNLFRERCGKRLRAKGLSGVEVDGL-----LDGSAWTPCRMTIFSAVMLTL--FRGV 1214
 DB 1490 IRLV-----HLIRSAWILQAPKESLRGFWMKKGSGEPSTLLMNTVMNAVITHCYDFRDL 1543
 QY 1215 KFAAFKGGDSL-LGSHLRFDASRLHNGERYKTKHLKVEVOKIVPYIGLVSAEQVYL- 1272
 DB 1544 QVAAFKGGDSLVLCESEYQSPGAAVLIAGCGLK--LKVDFRPIGLVAGVVVAPGIALP 1600
 QY 1273 DVPVSA 1278
 DB 1601 DVVRF 1606
 RESULT 12
 AAM80196
 ID AAM80196 standard; Protein; 1693 AA.
 XX
 AC AAM80196;
 XX
 DT 25-MAR-2003 (updated)
 DT 23-DEC-1998 (first entry)
 XX
 DE Protein encoded by ORF1 of ET-NANB (HEV) Burma strain DNA sequence.
 XX
 KM Enterically transmitted non A non B hepatitis virus; ET-NANB;
 KM Hepatitis E virus; HEV; Burma HEV isolate; vaccine;
 KM diagnostic probe.
 XX
 OS Non A non B Hepatitis virus.
 XX
 PN US5824649-A.
 XX
 PD 20-OCT-1998.
 XX
 PF 07-JUN-1995; 95US-0475807.
 XX
 PR 25-JUL-1994; 94US-0279823.
 PR 17-JUN-1988; 88US-0208997.
 PR 11-APR-1989; 89US-0336672.
 PR 16-JUN-1989; 89US-0367486.
 PR 13-OCT-1989; 89US-0420921.
 PR 05-APR-1990; 90US-0505888.
 XX
 PA (GENE-) GENELABS TECHNOLOGIES INC.
 XX
 PI Bradley DM, Fry KE, Krawczynski KZ, Reyes GR, Tam A;
 PI Yabough PO;
 XX
 DR MPI; 1998-582599/49.
 DR N-PSDB; AAV66321.
 XX
 FT Hepatitis E virus proteins - useful for diagnosis or vaccine
 FT production the virus
 XX
 PS Claim 22; Columns 57-66; 47bp; English.
 XX
 CC AAM80196-98 are encoded by the genome of the Burma strain of
 CC enterically transmitted non A non B hepatitis virus (ET-NANB)
 CC (hepatitis E virus (HEV)). The specification describes an isolated
 CC protein which is specifically immunoreactive with antibodies present
 CC in individuals infected with HEV and encoded by a sequence contained
 CC in an open reading frame (ORF) of an HEV genome. The genome has a
 CC sequence that is more than 70% identical to the ORF1 sequence from
 CC Burma HEV isolate. The protein is used as a vaccine and a diagnostic
 CC probe for ET-NANB.
 CC (Updated on 25-MAR-2003 to correct PR field.)
 XX
 SQ Sequence 1693 AA;

Query Match 5.4%; Score 479; DB 19; Length 1693;
 Best Local Similarity 21.2%; Pred. No. 1.9e-27;
 Matches 369; Conservative 186; Mismatches 577; Indels 610; Gaps 73;
 2 YAKATDVAAYVAADVAAYVAVYLQORAVKLDPAAPKALBETLHRLYYP--LRFKGTLPPT 59
 DB 6 FIKAPGTTTMEQALAAALANALANAVVVRPFLSHQDIBILINLMQRLVFRBEV--W 63
 QY 60 QHPILAGHVAEEVLANPARGNS-TVLEIGPSLSALKLKGADNAPVADYHG--TKYGT 117
 DB 64 NHPY-----ORVINEHELELYCARSGRCLEIGHNPRS--INDNNV-----VHRCFLRPVG 112
 QY 118 RDGSRHITALE-----SRVATGRPEFKADSLANGIASRFTCYDGVGSCAFKSVG 170
 DB 113 RDVGRWYTAFTPRGANCRSALRGLP-----AADRTYCLDGFSGCNFPAETG 160
 QY 171 IANSHLYDVTLEBIANAPEHNGLMVPAFMMPEELLYMDVNVAAELGYRFHVEBPMAV 230
 DB 161 IALYSLHDMSPSDVAEAMFRHGMTRLYALHLPPREVLL-----PRTGY 203
 QY 231 KDCAPFGGDLRLHPELDFINESQBRRIERLAAGSYSRRAVIFSGDDMDGAYLHD--- 287
 DB 204 RTAGY---LLIH-----DGRHV-----VVTYEGDTSAG--YNHDSN 235
 QY 288 FHTWLAVLVNRYPTPPGFSLHTEVQRHSGSIELRTTRAP-----PDRMLAVVR 339
 DB 236 LRSM-----ITTKVTGDHPLVIERVRAIGHFVLLTLAABSPMPVVPFRSTEVYVR 290
 QY 340 TSQGLCRIPNIFYVADASGEHKITLTSOKVMNLNFMQTRPKELVDMLVLSFAPAR 399
 DB 291 SIFPGGTPLSF--PTSCSTSTFHAVPAHIMDLMEFGATLDQAC--CSRLTYTLRGI 347
 QY 400 LRAIVASEVTESGWNISPADLVRTVSVLY-----LHIERRAAVAV 443
 DB 348 SYKTVGLVLANEGMNASBDALTAIVITTAAYLTICORYLRTQAIKGRRLERHAQFI 407
 QY 444 -----KXAKDVVG--ETSFWSLKHVLS----- 466
 DB 408 TRLYSMTPEKSGRDYIPGRQLEFYACRRWLASGPHLDPRVLVDESAPCHCTAIRKAL 467
 QY 467 -----CC-----GLNLKGTDVVFTKR 483
 DB 468 SKFCCEPKMKLQGECECTCLQPAEAGVGOGHNEAYEGSDVDVPASALSIDIGSYVVPETA 527
 QY 484 VDKRYVHSL-----GDIIICVRL-----SP 504
 DB 528 LQPLVQALDLPAEIVABAGRLTATVKVSQVYGRIDCELTGKNTFRTSFVYGALETNGP 587
 QY 505 EQ-----VGFLPSRVV----- 515
 DB 588 ERHNLSPDASQSTMAAGPFLITVAASAAGLEVRYVAAGLDHRVAVPAPVSPRSPAGEVTA 647
 QY 516 -----PARVPHD----- 522
 DB 648 PCSALYRFNREAGRHSLIGNIMFHPBGLIGLPAPFSQHVWESANPPCGESTLYTRTWE 707
 QY 523 -----RELEVLVRAG-----CYNERPVSTPVEBPQG----- 551
 DB 708 VDAVSSPARPDIGFMSSEPSIFSRATPTLAAPLPAPADPSPSAPALABAPAGATAGA 767
 QY 552 -----PDADLMAT----- 560
 DB 768 PAIHQTAHRHRLFTYPPDSKVPAAGSLFBSTCTMLVNASVNDHRPGGGLCHAFYORYPA 827
 QY 561 -----AASL-----ATLDAG-----LN 576
 DB 828 SFDAASTVMDGAAYTLTRPIIHAVAPDVRLEHNRKLEAAABRFGCSRLGTAAYPILG 887
 QY 577 TDVKK-----LKIT 585
 DB 888 TGIYQVPIGSPFDMABRNHRPGDELVLPLAARFPAANRPTRPYLITTEDVARTANLAIE 947
 QY 586 LBNL-----KITDGLTLPVRLGLENYE--GPRSGKGTGTLIALBAAGKALVYAPREL 639

```

DB 948 LDSATDVGRACGCKRTP--GVVQYQFTGVPSCGSKRSITQA-----DVVVVVPTREL 1000
OY 640 REAMDRRIKPSASATQHVAAIIL--RPAIAGAPATVVIDECFMPVLVVAIVHASP 697
DB 1001 RAAWRRR--GFAAFPHTAARVTOGR-----VIDEAPSLP-PHILLH-MQR 1045
OY 698 SRRIVLVGDVHOIGFIDFQGTSTANMPVLVDVVKQCRRTFFNOTKCRPADVVAATFFQSLY 757
DB 1046 AATVHLGDPNQIPALDFEHAGL-VPALRPDLGP--TSMWHTHRPAPV--CELIRGAY 1100
OY 758 PCSTTSGCVASISHAPDRYNSQQTLCFTOEKSRHAGAGMTHEAGRTFASVILH 817
DB 1101 PMITOTTSRVKSLFMCBP-----AVGQKLVFTQAKPAN--PGSVTHVHQGATYETTTII 1154
OY 818 YNGSTAEOKLAEKS--HLLVGIITRTHNLVIRDPGTGIERQLNSAKAEVFTDIPARLE 875
DB 1155 ---ATADANGLIGSSAAHAI VALTRHTEKCVIIDARG-LLREVGIS-DAIVNNFLAGE 1209
OY 876 ITTVKPSSEVQRN-----EVMATIPQSATPHGAIHLKRNQGDQDCCVALAKTGYEV 930
DB 1210 IGHORPSVIPRGNPPANVDTLAAPPSCQI--SAFHQLABEIGHRP-VVAAVLPCCPEL 1266
OY 931 FCGRAKINVELAEPDTPKPHRAFO-EGYQWVYVTVASNKQALOTLLSRYTKRSADLP 989
DB 1267 EGGLLYLPQELTTTCSV-----TFELTIVHCRMAAPSORKAVLSTLVGRIGRT--KL 1319
OY 990 HEAK-EDVVRMLNSLDRHMDWTVEDARDRAVETOLKFTORGIVEDLLEPD--DPYIR 1046
DB 1320 YNASHSDVR--DSLARFIPAIQPVQVTTCELYEVAWEKQDQSAVLIEDLCNRDVS 1376
OY 1047 DIDFLMKTQOK-VSPKPIITGVYGGGIAHSSKSLNVLNARILEE-ILRTGSRVYRS 1104
DB 1377 RTTFPKDKNCKTTGTGIIHAGKVGQGISAMSKFCLFGPWFPAIKALIALLPQSVFYG 1436
OY 1105 NGLPDEE-EAMLEAKINOVPHATFVSAMTEPDTNHNNTSELLFAALIRIGTPAAAN 1163
DB 1437 DAFDDVFSAAVAAAASV-----FENDSEBDSIQNNFSLGLECAINEBCMPWLIR 1491
OY 1164 LFRERCKRTLRPAKIGSVEVDGL-----LDSCAATPCRNITFSAVMTL--FRGVK 1216
DB 1492 LV-----HLIRSAWILQAPKSLRGFWKHGSEPGTLLMNTVMNNAVITHCDFRDPQV 1545
OY 1217 AAFKGDGSL-LGSHYLRFDASRLHNGERYKTKHLKVEVQKIVPYIGLLVSAEQVTL-DP 1274
DB 1546 AAFKGDGSLVLCSEYRQSPGAAVLIAGCGLK--LKVDPRPIGLVAGVVAVPGLALPDV 1602
OY 1275 VRSR 1278
DB 1603 VRPA 1606

```

```

RESULT 13
AAB24119
ID AAB24119 standard; Protein; 1693 AA.
XX
XX AAB24119;
XX
XX 29-JAN-2001 (first entry)
XX
XX Hepatitis E virus Burma strain ORF1 protein sequence SEQ ID NO:7.
XX
XX Hepatitis E virus; HEV; non-A/non-B viral hepatitis; viral protein;
XX immunological; diagnosis; hepatitis; infection; identification;
XX detection; immunoreactive; hepatotropic; antiinflammatory; vitruclide;
XX vaccine; antiviral; antigenic; antibody; antigen.
XX
XX Hepatitis E virus.
XX
XX US6120988-A.
XX
XX 19-SEP-2000.
XX

```

```

PF 07-JUN-1995; 95US-0478507.
XX
XX 25-JUL-1994; 94US-0279823.
PR 05-APR-1991; 91US-0681078.
PR 11-JUN-1988; 88US-0208997.
PR 11-APR-1989; 89US-033672.
PR 16-JUN-1989; 89US-0367486.
PR 13-OCT-1989; 89US-0420921.
PR 05-APR-1990; 90US-0505888.
XX
XX (GENE-) GENELABS TECHNOLOGIES INC.
XX (USGO ) US GOVERNMENT.
XX
XX Yardough PO, Krawczynski KZ, Fry KE, Bradley DW, Tam A, Reyes GR;
XX N-PSDB; AAA55259.
XX
XX WPI; 2000-593712/56.
XX
XX Identifying recombinant antigen immunoreactive with antibody induced by
XX hepatitis E virus (HEV), for detecting HEV infection, comprises
XX immunoreacting a polypeptide from an HEV genome with an HEV-positive
XX antiserum.
XX
XX Claim 6; Column 55-64; 46pp; English.
XX
XX The present invention describes a method for identifying a recombinant
XX antigen immunoreactive with a hepatitis E virus (HEV)-induced antibody.
XX The method comprises producing a polypeptide derived from an HEV genome,
XX immunoreacting the polypeptide with an HEV-positive antiserum and
XX selecting the polypeptide as a recombinant antigen if the polypeptide
XX reacts with the HEV-positive antiserum. The method is useful for
XX identifying recombinant antigen immunoreactive with antibody induced by
XX HEV. The enterically transmitted non-A/non-B hepatitis (ET-NANB) (also
XX known as HEV)-specific fragments are useful for identifying ET-NANB
XX derived cDNAs, which contain additional sequence information, as primers
XX for detecting ET-NANB viral genomic material in a patient sample, for
XX the synthesis of polypeptides for use in immunoassays, and for
XX identifying similar antigenic regions encoded by related viral strains,
XX e.g. Burmese strain. The antigens are especially useful in the
XX preparation of vaccine against ET-NANB infection. These antigens may
XX further be used to prepare antibodies to ET-NANB virus particles for
XX use directly as antiviral agents, and to produce antiserum designed for
XX pre- or post-exposure prophylaxis. The present sequence represents a
XX specifically claimed HEV Burma strain protein sequence for use in the
XX present invention.
XX
XX Sequence 1693 AA;
XX
XX Query Match 5.4%; Score 479; DB 21; Length 1693;
XX Best Local Similarity 21.2%; Pred. No. 1.9e-27;
XX Matches 369; Conservative 188; Mismatches 577; Indels 610; Gaps 73;
XX
OY 2 YAKATDVAVVAADAVYANVLQORAVKIDPAPPLKALFTLRLYYP--LRPKGTLPTPT 59
DB 6 FIKAPGITTAIQALAAANSAALANAVVRPPLSHQOIEILINLMQPRQVLRPRVF--W 63
OY 60 QHPITLAGHGVAEVLNFAGRS--TVEIGPSLSALQHGAPNAPVADYHGC-TKYGT 117
DB 64 NHPI-----QRVINHELELYCRARSGRCLEIGHAPRS--INDNPNV-----VRCFLRPVG 112
OY 118 RDGSRHITALE-----SRSAVATGRPEPKADASLLANGIASTRPCVDGVGSCAPSRVG 170
DB 113 RDVQRYNTAPTPTGPAANCERSLRGLP-----AADRTYCLDPSGCGNFPALRTG 160
OY 171 IANHSLYDVTLELANAFENHGLHWRAFPHNPEBLLYNDVNVNAELGYRFHVBEPMAV 230
DB 161 IALYSLHWSPSDVAEAMFRHGTRLYALHLPPEVTL-----PPGYT 203
OY 231 KOCAFQGDRLRHPPELDFINESQERRIERLARQSYSRRAVIFSGDDMDGAYLHD--- 287
DB 204 RTASY--LLIH-----DGRVV-----VVTYEDDTAG--YNHDSVN 235
OY 288 FHTWLAIVLVRYPTPFQPSLHIBVQRHGGSSIBLRITRAP-----PGDRMLLVVR 339

```

```

Db 236 LRSW-----IRTKYTGDPHPIVERVRAIGCHPULLLTAAPSPSPMPYPRVPRSTREYVR 290
Qy 340 TSQGLCRIPNIFYVADASTGTEHKTLTTSQHKVMMLNFMQTRPEKELVDMVLMSFAPAR 399
Db 291 SIFGCGSPSLF--PTSCSTKSTFHAVPAHIMDRIMLFGATLDDQAF-C SRLMTYLGDI 347
Qy 400 LRAIVVASEVTESSNNISPADLVRTVSLVY-----LHIIIRRAAAV 443
Db 348 SYKVVGLTVANEGNASEDALTAIVITAAVLTICHORYLRTQAIKSGMRRLREHAQKFI 407
Qy 444 -----KTKADDFG-ETSPFESLKVLS----- 466
Db 408 TRIVSWLPEKSGRDYIPGRQLEFYAQCRRKLSAGHLDPRVLVPDESAPCHRTAIRKL 467
Qy 467 ---CC-----GLRLNKTGDDVFTKR 483
Db 468 SKFCFEMKMGQECTFLQPAEAGAVDQGDHNEAYGSDVDPAESAISDLSGYVPGTA 527
Qy 484 VVDKRVHSL-----GDIIICDVR-----SP 504
Db 528 LQPLTQALDLPRAEIVARAGRLTATVKSQVDRIDCETLLGNKTRFRTSFDGAVLETNGP 587
Qy 505 EQ-----VGFLPSRPV----- 515
Db 588 ERHNLSPDASOSTMAAGPESLTYAASAGLEVRVYAAGLDHRAVAPGVSPASAGEVTA 647
Qy 516 -----PARVFD----- 522
Db 648 FCSALYRREARQASHLIGNLWFHPEGLGLPAPSPGHVMSANPCGSESTLYRTWSE 707
Qy 523 -----RELEVUREAG-----CYNERVBPSTPRPEEG----- 551
Db 708 VDAVSSPARPDGFMSEPSIPSRATPTLAAPLPPAPPPSPAPALAPASAGATGA 767
Qy 552 -----PDADLWHAT----- 560
Db 768 PAITHQTRARHRLTTYPDGSKVFGASLFESTICTLVNANSVDHRRPGGLCHAFFQRYPA 827
Qy 561 ---AASL-----PEYR-----ATLQAG-----LN 576
Db 828 SFDASFWMRDGAAYTLTPRIHAVAADVRLHNPKLEAAYRETSRLGTAAYPLG 887
Qy 577 TDVKQ-----LKT 585
Db 888 TGIYQVPIGPSEFAMERHRRPGDELYLPELAARMEANRPTTLITTEDVARTANLAE 947
Qy 586 LENAL---KTIDGLTSPVRLGEMYE---GPPSGKGTGLIAALEAGSKALVYAPTREL 639
Db 948 LDSATDVGACAGCAYTP--GVVQYQFTAGVFGSGKSNITQA-----DVDVVVPTRRL 1000
Qy 640 REAMDRIRKPPSASATOHVALAIL--RRATAGAFATVVIDECMFPLVVAIVHALSP 697
Db 1001 RNAMRR---GFAAPTPTTAARVTOGR-----VIDEAPSLP--PHILLH--MQR 1045
Qy 698 SSRIVLVGDVHOIGTIDFGTSANMPLVADVVKQCRRTFNQTKCPRADVAATTFQSLY 757
Db 1046 AATVALLDQDPNOIPRAIDEPHAGL--VPAIRPDIGP--TSMWHTHMPADV--CELRIGAY 1100
Qy 758 PCCTTSGCVASISHVADPYRNSAQOTLCFTQEBKSRHGAEGAMVVHEAGRTFASVLH 817
Db 1101 PMIQTTSKVLSLFGEP---AVGQKLVFTQAKPAN--PGSVVHEHQAGATYVETITII 1154
Qy 818 YNGSTAQOKLAESK--HLVGTIRHTNMLYIRDPGTDIRQJNHSAKAEVFTDIPAPLE 875
Db 1155 ---ATADARGLIQSRAHAIVALTHTKECVIIDAQ--LLREVGIS--DAIVNPFIDGE 1209
Qy 876 ITTVPSSEVQRN-----EWATTIPQSTPHGAIHLKRNKNGDDCCCVALATGYEV 930
Db 1210 IGHORPSVI PRGNPDANVDTLAAPPSCOI--SAHQIQAELBEGHRP--PVAAVLPPCEPL 1266
Qy 931 FGGRAKINVELAEPDAPTKPHRAFO--EGVQWKVNNASKHQALQTLISRYKRSADPL 989

```

```

Db 1267 BQGLLYLPQBLTTCDSVY-----TFELNDIVHCRMAAPSGRAVSLTVGRYGRF---KL 1319
Qy 990 HEAK--EDVKRMKNSIDRMWMTTVEDADRAVFFETQLKFTQRGTEVLEDPD--DPYIR 1046
Db 1320 YNASHSDVR---DSLARFI PAIGPVYTTCELYELVEAMVEGQGSVLELDLCNRDVS 1376
Qy 1047 DIDFLMKTOOK--VSPKPIINTGKVGQGIHAHSKSLNFPVLAAMRIIEB--ILRGSTYRVS 1104
Db 1377 RITFPQCKNFTTGETTIAHGKVGQGISAMSKTFCALGPMFRALKAIALLPQGVFYG 1436
Qy 1105 NGLPDEE--EAMLEAKINQVPHATFVSADMTFEPDTHANNTESELPALLERIGTPAAAVN 1163
Db 1437 DAFDDTVSAVAVAAKAMV-----FENDPSEFSTQNNFSLGLSCALMEBCGMQMLIR 1491
Qy 1164 LFRERCGRTIRAKGLSGVEVDGL-----LDSGAATPCRNITFSAAVMLTL--FRGYKF 1216
Db 1492 LY-----HLIRSAWILQAPKESIRGFMKKSGSEGTLLMTVMMAVITHCYDRDFQV 1545
Qy 1217 AAFKGDSTL--LCGSHYLRFDASRLHMGERYTKHLXVEVOKTVPYIGLLVSAEQVYL--DP 1274
Db 1546 AAFKGDSDIVLCSBYRQSPGAVALIAGCGLK---LKVDPRPIGLVAGVVAAGALPDPV 1602
Qy 1275 VRSA 1278
Db 1603 VRFA 1606

RESULT 14
AAB62522
ID AAB62522 standard; Protein; 1693 AA.
XX
AAB62522;
AC
XX
DT 23-JUN-2001 (first entry)
XX
DE HEV--Burma strain viral protein ORF1.
XX
KM HEV, enterically transmitted nonA/nonB viral hepatitis agent; viral;
XX
KW ET-NANB; infection; vaccine.
XX
OS Hepatitis B virus.
XX
PN US6229005-B1.
XX
PD 08-MAY-2001.
XX
PF 03-AUG-1998; 98US-0128275.
XX
PR 05-APR-1991; 91US-0681078.
PR 25-JUL-1994; 94US-0279823.
PR 17-JUN-1988; 88US-0208997.
PR 11-APR-1989; 89US-0336672.
PR 16-JUN-1989; 89US-0367486.
PR 13-OCT-1989; 89US-0420921.
PR 05-APR-1990; 90US-0505888.
XX
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA (GENE-) GENELABS TECHNOLOGIES INC.
XX
Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam AW, Fry KE;
WPI, 2001-342705/36.
DR N-PSDB; AAF83485.
XX
New DNA sequences of enterically transmitted non-A/non-B (ET-NANB)
PT hepatitis viral agent, useful in diagnosing infection by an enterically
PT transmitted agent (e.g. ET-NANB virus), as well as in vaccine
PT production.
XX
PS Discloure; Columns 55-64; 45pp; English.
XX
CC The invention relates to an isolated DNA comprising the genome of an
enterically transmitted nonA/nonB (ET-NANB) viral hepatitis agent (also

```

referred as HEV). The DNA sequences or their fragments are useful in preparing ER-NANB viral proteins and as probes for virus detection. These are particularly useful in diagnosing infection by an enterically transmitted agent (e.g. ER-NANB virus), as well as in vaccine production. The present sequence represents a ORF1 protein encoded by a ER-NANB viral DNA sequence from HEV-Burma strain.

Sequence 1693 AA:

Query Match 5.4%; Score 479; DB 22; Length 1693;
Best Local Similarity 21.2%; Pred. No. 1,9e-27;
Matches 369; Conservative 188; Mismatches 577; Indels 610; Gaps 73;

2 YAKATVAVRYAADAANVAVNLOQRAVKLDPAAPLKALFTLRLVLP--LRPAGTILPPT 59
6 FIAAPGITTAEIOAALANANALANAVVPRFLSHOQIEILINIMOPROQLVFRPEVF--W 63
60 QHPILAGHORVABEVVHNFAGRS--TVLEIGPSLSHSAKLHGAPNAPVADYHGC--TKYGT 117
64 NHPI---QVHINELVELYCRARSGRCLEIGHAPRS---INDNPNV---VHRCFLRPVG 112
118 RDSRRIITALE-----SRSVATGRPEFKADASLLANGIASRTFCVGVGSCAFKSRVG 170
113 RDVQRVYTAPTRGPAANCRRSALRGLP-----AADRTYCLDGFSGCFNFAETG 160
171 IANSLYDVTLEELANAFENHGLHVRAPFHMPEBELLYMDNVNAELGYRFHVIIEPMAY 230
161 IALYSIDHMSPSDVAEMFPHGMYRLAALHLPRPEVL-----PGTY 203
231 KDAFQGGDLRHFPELDFINESQERIERLAARGSRRRAVIFSGDDMDGDAYLHD--- 287
204 RTASY---LHIIH-----DGRRV-----VVTYEGDTSAG--YNDHVS 225
288 FHTMLAVLVNRPTEPFGLHIEVORRHSSIELRTTRAP-----PGDMILAVVPR 339
236 LRHM-----IRTKVTGDHVLIERVPAIGHFVLLITAPERSPMYVYPPSTEVYVR 230
340 TSGCLGRIPNIFYVADASGTEHKTILTSQKVNMLNFMQTRPEKELVDMTILMSFARAR 399
291 SIFPGGTPLSL--PFSCTKSTFHAHPAHIMRLMFGATLDDQAF--CSRLMTYLRGI 347
400 LRIVASVETSSWNSPADLVRTVSLV-----LHIIERRRAVAV 443
348 SYKVTGTVANEGWNASBDALTAVITAAVLTICHQRYLTQALISKMRLEBHAQFI 407
444 -----KTAQDVFG--ETSFWESLKHVLS-----GLRLKGTIDVFTKR 466
408 TRLYSWLFKESGRDYIPGRQLEFYAQCRLSAGFHLDPVLPVDESAFGCHRTAIRKAL 467
467 ---CC-----GLRLKGTIDVFTKR 483
468 SKFCCEFMKWLGOECTCFLPAEGAVGDQHDNEAEGSDVDPAESALDISGSVYVVGTA 527
484 VDKRYVHSL-----GLICVRL-----SP 504
528 LQPLDALDLPRAIIVARAGLTATVKSQVDGRIDCETLLGNKTFRTSFVDGAVLETNGP 587
505 EQ-----VGFLPSRPV----- 515
588 ERINLSFSDASQTMAGPFLTYAASAGLEVRYVAAGLDHRVAFAGVSPRSGVTA 647
516 -----PARVPHD----- 522
648 FCSALYFNRBAORHSLIGNLMFHBGLIGLFAFPFSGHGWESANPFCGSESTLYTRTWE 707
523 -----RELEVLREAG-----CYNERPVSTPVEEPOG----- 551
708 VDAVSSPAPDLPGLFMSPEPSIPSPAAFTTLAALPPEPPAPDPSPPSAPALAPASGATAGA 767
552 -----FDADLMHAT----- 560
768 PAITHQTAHRRLFTYPPDGSKVFPAGSLFESTCTMVLVNASVDHRRPGGGLCHAFYQRYPA 827

561 ---AAL-----PBYR-----ATLQAG-----LN 576
828 SPDAASFVNRDGAAYTLTPRPIIAVADPYRLBNPKLEAYRBTCSRLGTAAVPLLG 887
577 TDVKQ-----LKIT 585
888 TGIYQVPIGSPDAMERANRPDDELYPELAAWFMFANPRTPTLTITTEDVARTANLAI 947
586 LENAL---KITDGLTSPVGLMEYE---GPPSGKTGTLIALBAAGKALVYAPTRRL 639
948 LBSATDVGAACAGCRVTP--GVVQYFTAGVPPSGSRSTQA-----DVDVVVVTREL 1000
640 REAMDRIRPPSASAGVVALIL--RAATBAPAPATVYIDECFMPFLVYVAIVALS 697
1001 RNAMRR---GPAALFPHTAARTQGR-----VVIDEAPSLP--PHLLH--MOR 1045
698 SSRIVLGDVHNGIFIDPQGSANMPLVDVNVQCSRRATFNQKRCRADVATTPQSLY 757
1046 AATVHLIDPNQIPALIDFENAGL--VPAIRPDLP--TSMWVTHRPADV--CELIRGAY 1100
758 PGCTTSGCVASISHVAPDYRNSOAGTLCTQEBKSRHAGBAMTVHBAQGRTPASVILH 817
1101 PMIQTSRVLRLSLFWGER---AVGQKLVFTQAKKAN--PGSVTVHBAQATYETETII 1154
818 YNGSTAEQULAEKS--HLVGTTRTNILYTRDPGDIEROLNHSKAEVFTDIPAPLE 875
1155 ---ATDARGGLQSSRAHAIVALTRTEKCVIIDAAG--LIREGIS--DAIVNNFPLAGGE 1209
876 ITTVKPSBEVQRN-----EVMATIPQASATPHGAILHAKNPFQDQCCGVALLAKGYEV 930
1210 IGHQPSVLPKGNPDANDTILAAFPSCQI--SAFQQLAEGLHNP--VPVAAVLPPCPREL 1266
931 FGGRAKINELAEPPATPPKRAFO--EGVQWVVTNASNKGALQTLSTRYKRSADLP 989
1267 EGGLVLPQGLTTCDSVV---TFELTDIVHCSMAAPSGRKAVLSTLVGRYGGRT---KL 1319
990 HEAK--EDVKMLNSLDHMDWTEDARBAVFEOLKTKQKGTVEDLLEP--DPYIR 1046
1320 YNASHSDVR---DSLARFPAIGPVVTCLELVELEVAWEKQDQSAVLELDLCNRDVS 1376
1047 DIDFLMKTOOK--VSPKINTGKVGQIAGHSKSLNVLAMRILBEE--ILRTGSRTRYRS 1104
1377 RITFPQDCNKFTTGTTIAGKVGQISMSKTPCLPGWFAIRKAILALLPOGVFTG 1436
1105 NGLPDEE--EAMLEAKINOVPHATFVSADWTEBDTANNYSBELFALLERIGTPAAAVN 1163
1437 DAFDDTVFSAVAALAAASNV-----FENDFSEBDSIQNNPFSGLBGLMBEGMPQWLIR 1491
1164 LPRBRCGKTLRAKGLGSVVDGL-----LDSGAMTPPCANTIPSAVNLTL--FGVYR 1216
1492 LY-----HLIRSAMLIQAPKESLRGFWKKGSGBPOTLLMNTVMNNAVITHCYDFRDPV 1545
1217 AAFKGDLSL--LGSHTLRFDPASRLHNGERTKYLKUYEOKIYPIGLVSAEQVVL--DP 1274
1546 AAFKGDLSYLCSERYQSPGAAVLGAGCLK--LVNDRPRLGLYGVVAVAPBLGALPDV 1602
1275 VRGA 1278
1603 VRPA 1606

RESULT 15
AA015698
ID AA015698 standard; Protein; 1693 AA.

XX AA015698;

XX 08-NOV-2002 (first entry)

XX Hepatitis B virus (Burma strain) ORF1 protein.

XX HEV; enterically transmitted nonA/nonB hepatitis viral agent;
XX Burma strain; bile; ORF1.

```

XX OS Hepatitis E virus.
XX PN US6379891-B1.
XX PD 30-APR-2002.
XX PF 19-APR-2000; 2000US-0553427.
XX PR 25-JUL-1994; 94US-0279823.
XX PR 07-JUN-1995; 95US-0478507.
XX PR 15-APR-1991; 91US-0681078.
XX PR 17-JUN-1988; 88US-0208997.
XX PR 11-APR-1989; 89US-0336672.
XX PR 16-JUN-1989; 89US-0367486.
XX PR 13-OCT-1989; 89US-0420921.
XX PR 05-APR-1990; 90US-0505888.
XX PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.
XX PA (GENE-) GENELABS TECHNOLOGIES INC.
XX PI Reyes GR, Yarbough PO, Bradley DW, Krawczynski KZ, Tam A, Fry KE;
XX DR WPI; 2002-517277/55.
XX DR N-PSDB; AAL50386.
XX PT Detecting the presence of enterically transmitted nona/nonb hepatitis
XX PT viral (HEV) agents in bile samples from infected humans and monkeys
XX PT using polymerase chain reaction -
XX PS Disclosure; Column 13-20; 61pp; English.
XX CC The invention comprises a method for detecting the presence of
XX CC enterically transmitted nona/nonb hepatitis viral (HEV) agents in a
XX CC sample and isolating HEV agents or nucleic acid fragments produced by the
XX CC agent. The method utilizes PCR - using bile from a human or cynomolgus
XX CC monkey actively infected with HEV as a source of the agent. The method of
XX CC the invention is used for detecting the presence of a viral agent in a
XX CC sample of cultured cells infected with the agent and isolating
XX CC enterically transmitted nona/nonb HEV agents or nucleic acid fragments
XX CC produced by the agent. The present amino acid sequence represents the
XX CC protein encoded by open reading frame 1 (ORF1) of a DNA sequence isolated
XX CC from a Burmese strain of the Hepatitis E virus.
XX CC
XX Sequence 1693 AA;
XX
Query Match 5.4%; Score 479; DB 23; Length 1693;
Best Local Similarity 21.2%; Pred. No. 1.9e-27;
Matches 369; Conservative 188; Mismatches 577; Indels 610; Gaps 73;
QY 2 YAKATDVAVVAADVAAYANVLOQRAVLDPAFLKALETLRLYYP--LRFGGGLPPT 59
DB 6 FIKAPGTTTAEQALAAALANAVVVRFLSHQOIEILINLMQRPQLVFRREV--W 63
QY 60 QHPLAGHQVAAEVLNHFARGRS- TVEIGPSLHSAIKLHGAPNAVADYHG-C- TKXGT 117
DB 64 NHPI-----QRIYHNELELYCRASGRCLIEIGAPHS--INDNPV-----VHRCFLRPVG 112
QY 118 RDGSRHITALE-----SRVATGRPREFKADSLANGIASRTFCVDCVGS-CAFKSVG 170
DB 113 RDVGRWYTAPTRGPANCRSALNGLP-----ADRTYCLDGFSGCNFPATG 160
QY 171 IANSLYDVTLAEILANAFENHGLHVRFAFMPEELLYMDVVVAALGYRPHVIEEPMAV 230
DB 161 IALSLHMSPSDVAEMFRGQTRLYAALHLPREVL-----PQTY 203
QY 231 KDCAPFGGDLRLHPELDFINESQERRIERLAAGSYSRRAVITSGDDMDGDAYLHD-- 287
DB 204 RTASY-----LTIH-----DGRV-----VVTYEGDTSAG--YNHVDSN 235
QY 288 FHTMLAVLYLVNYPTRPFGSLHIEVQRHSGSIELRITRAP-----PGDRMLAVVPR 339
DB 236 LRSM-----IKTKVTGDHPLVIERVRAIGCHFVLLTLAABPSPMPYVYPRSTEVVR 290

```

```

QY 340 TSGCLCRIPNIFYAASGTEHKTILTSQKVMNLNFMOTREKELVDMTVMLSFAR 399
DB 291 SIFPGGTFSLF--PTSCSTKSTFHAVPAHIMWLMFGATLDQAPC--CSRLMTYRGI 347
QY 400 LRAIVASEVTBSSWNISPADLVRTVSLV-----LHIERRAAV 443
DB 348 SYKTVGTIVANESMNASBDALTAVITAYLITICHQYLTQALISKMRRLERBAQKFI 407
QY 444 -----KTADDVVG-ETSPWESLKHVLS----- 466
DB 408 TRLSWMLFEKSGRDIYIGRLBFYAQCRWLSAGFHLDPRLVLPDESAPCHRAIRKAL 467
QY 467 ---CC-----GLRNLKGTDVFTKR 483
DB 468 SKPCCFMKWLQGCETCELPABAGVGDQHDNEAYGSDVDPAESAISDISGYVVGTA 527
QY 484 VVDKYRHSI-----GDICDVR-----SP 504
DB 528 LQPLYLADLPALIVARAGRLTATVKVSYVDGRIDCETLLGKTKFRTSPVDGAVLETNGP 587
QY 505 EQ-----VGLPSPRP----- 515
DB 588 ERHNLSPDASGSTWAGPFLTYTAASAAGLEVRYVAGLDHRAVAFAGVSPRSAPGETA 647
QY 516 -----PARVPHD----- 522
DB 648 FCSALYFNBEOAHSILGNLMFHEBGLIGLPAPFSPGHWSANPFGESTLYRTWSE 707
QY 523 -----REELBYLREAG-----CYNERPVSTPVEBPG----- 551
DB 708 VDAVSSPARPDGLGWSPPSIPSRATPTLAAPLPAPDPSPPSAPALASPASGATAGA 767
QY 552 -----FDADLMAT----- 560
DB 768 PAITHQABRRRLFTYPPDSKVPAGSLFESTCWLNVASVNDHRPGGLCHAFYQRYPA 827
QY 561 ---AASL-----PEYR-----ATLQAG-----LN 576
DB 828 SPDAASFVMBDAAYLTTPRTIHAVAPDRLHNPKRLBAAYRETCRSIGTAAPYLIG 887
QY 577 TDVQ----- 585
DB 888 TGIYQVPIGSPDAMERNHRPGDELYLPDLAARNFEANRPTRPTLTTEDEVARTANLAIS 947
QY 586 LENAL---KITDGLTSPVRLKEMY---GPPSGKGTGLIALBAAGKALVYAPTEEL 639
DB 948 LDSATDVGRACAGGRVTP--GVVOYQFTAGVPSGKGRSITQA-----DVDVVVPTREL 1000
QY 640 REAMDRRIKPPSASATQHVALAIL--RRATBAGAPFATVIDSCFMPPLVYVAIVHALSP 697
DB 1001 RNAMRRR---GFAFTDHTAARVTOGRR-----VIDEAPSLP--PHLLILH--NQR 1045
QY 698 SSRIVLVGDVHQIPIFPGTSANMPLVRODVVQCRRTFNQTCRCAPDVATTPFQSLY 757
DB 1046 AATVHLGDPQIPIAIDFENAGL--VPAIRDLGP--TSMMVITRWMPADV--CELIRAY 1100
QY 758 PCGTTGSCVASISHVAPDRNSOQTLCTQBEKSRHAGAGMTVHEAOCRTFASVILH 817
DB 1101 PMITGTSRLVLSLWGP-----AVGQCLVFTQAAKPPAN--PGSTVHEAQAATTETTTII 1154
QY 818 YNGSTAQKLLAEKS--HLVGTIRATNHLIYRPTGDIERQNLHSAKAVFTDIPAPLE 875
DB 1155 ---ATAADARGLIGSSRAHAIVVALFRHTEKCYIIDAPG--LIREVGIS--DAIVNPFLLAGE 1209
QY 876 ITTYKSEBEVGRN-----EVMATIPPOSATPHGAIHLIRKRFQGPQDCYALAKTGV 930
DB 1210 IGHORPSVTPRGNDANDVDTLAAPFSCQI--SAFHQLAEBELGRP--VPVAALVPCPEL 1266
QY 931 FGRAKINVELAEBDATPKPHRAFO--EGVGWVKVTNSNKHQALQTLSTRYTKSADLP 989
DB 1267 EGGLELIPQELITCDVV-----TELDIDYHCRMAASQRAVUSTLVGRIGRT--KL 1319

```

```

Qy 990 HEAK-EDVCRMINSIDRHMWMTVEDARDRAVEFETQKFTORGTEVDELLBPD--DPYIR 1046
Db 1320 VNAHSDVR---DSLARFIPAIGPVQVTCETELVEAMVEKQDGSATLEEDLCNRDVS 1376
Qy 1047 DIDFLMKTQOK-VSPRPINTGKVGOGIAASHSKSLNFWLAAMIRILEE-ILRTGSRTVRY 1104
Db 1377 RITFQKDCNKFYTTGETIAHGKVGQGISAMSKTFCALFGPWFRATEKAILLALPQVFPYG 1436
Qy 1105 NGLPDEE-EAMLEAKINOVPHATFVSADWTEPDTAANTSELFPALLERIGTPAAAYN 1163
Db 1437 DAFDDTVFSAVAVAASAVV-----FENDFSEFDSYNNFSLGLECAIMEBCGMPQMLIR 1491
Qy 1164 LFRERCGRKTLRAKGLGSEVVDGL-----LDGSAWTPCNRITFSAVM/TL--FRGVKF 1216
Db 1492 LY-----HLIRSAWILQAPKESLRGFWKXHSSEPGTLLMNTVMMAVITHCYDFRDPQV 1545
Qy 1217 AAFKGDLSL-LCGSHYLRFDPASRLHNGERYKTHLKVVEVOKIVPYIGLVSABQVVL-DP 1274
Db 1546 AAFKGDLSIVLCSEYRQSPGAATVLIAGCGLK---LKVDFRPIGLYAGVVVAPGLGALPDV 1602
Qy 1275 VRSA 1278
Db 1603 VRFA 1606

```

Search completed: January 15, 2004, 16:49:11
 Job time : 91.0574 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 16:26:38 ; Search time 34.0655 Seconds
(without alignments)
4810.480 Million cell updates/sec

Title: US-09-991-262-40

Perfect score: 8832
Sequence: 1 MYAKATDVAAVRYAAADVAYA.....LKRIKSTSDSVDPVKISKSA 1704

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 76:*

1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	542	6.1	1691	1 A44212	genome polypeptide
2	479	5.4	1693	1 MNMWH	genome polypeptide
3	323	3.7	1615	1 MNMWH	180K protein - tom
4	298.5	3.4	1597	2 S65053	genome polypeptide
5	298	3.4	1611	1 MNMWH	183K protein - pep
6	294	3.3	1694	2 UQ1896	hypothetical 191.1
7	291	3.3	1616	1 UQ2144	183K protein - tom
8	278	3.1	1615	1 MNMWH	183.3K protein - t
9	268.5	3.0	1608	1 MNMWH	183K protein - tom
10	265.5	3.0	1601	2 S48699	178K protein - tom
11	261.5	3.0	2115	2 S38480	nonstructural prote
12	253	2.9	1116	2 S26358	hypothetical prote
13	248.5	2.8	2205	1 MNMWH	nonstructural poly
14	242.5	2.7	1646	1 MNMWH	186K protein - cuc
15	242	2.7	1112	2 A49432	replicase 126K - o
16	231	2.6	1729	2 A49282	fusion protein 1a/
17	228.5	2.6	961	1 P1BVA	RNA 1a protein - b
18	214.5	2.4	1707	2 S01865	genome polypeptide
19	212	2.4	958	1 P1BVC	1a protein - compe
20	205	2.3	4776	2 E95206	cell wall surface
21	201	2.3	3175	1 RRVAV	genome polypeptide
22	196	2.2	1126	1 MNMWH	125K protein - alf
23	191.5	2.2	1224	2 T14007	microtubule-associ
24	189.5	2.1	290	2 C48547	nonstructural prote
25	188.5	2.1	2481	2 D90011	nonstructural prote
26	186.5	2.1	1367	2 S48478	glucan 1,4-alpha-g
27	184.5	2.1	1104	2 S59310	probable membrane
28	182	2.1	966	1 P1BVB	RNA 1a protein - b
29	181	2.0	1766	2 S03701	141K protein - pea

30	179.5	2.0	383	2 T46707	proteophosphoglyca
31	177	2.0	801	2 T39018	hypothetical prote
32	176.5	2.0	813	2 S70795	vea protein precu
33	175.5	2.0	879	2 S23006	shed acute-phase a
34	173	2.0	507	2 T44768	antifreeze glycop
35	173	2.0	786	2 T16509	hypothetical prote
36	170.5	1.9	725	2 A41258	a-egglectin core
37	170	1.9	1139	1 PAVB8	alpha-a protein -
38	169	1.9	1823	2 S28974	vitellogenin precu
39	168.5	1.9	993	2 U40074	RNA 1 protein - cu
40	167.5	1.9	1490	2 S72351	nonstructural poly
41	166.5	1.9	993	2 E71392	ATP-dependent RNA
42	166	1.9	660	2 JMO067	chitinase (EC 3.2.
43	165.5	1.9	993	1 P1VXM	RNA 1 protein - cu
44	165.5	1.9	2774	2 A43358	microtubule-associ
45	164.5	1.9	839	2 F75518	hypothetical prote

ALIGNMENTS

RESULT 1

A44212
genome polypeptide - hepatitis B virus (strain Mexico)
N:Contains: RNA-directed RNA polymerase (EC 2.7.7.46)
C:Species: hepatitis B virus
C>Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 19-Jan-2001
R:Huang, C.C.; Nguyen, D.; Fernandez, J.; Yun, K.Y.; Fry, K.E.; Bradley, D.W.; Tam, A.
Virology 191, 550-558, 1992
A>Title: Molecular cloning and sequencing of the Mexico isolate of hepatitis B virus (
A:Reference number: A44212; MUID:93079857; PMID:1448913
A:Accession: A44212
A:Molecule type: genomic RNA
A:Residues: 1-1691 <HDA>
A:Cross-references: GB:M74506; NID:9330017; PID:AAA5730.1; PID:9330018
R:Fry, K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Piatak, M.; Fe
Virus Genes 6, 173-185, 1992
A>Title: Hepatitis B virus (HBV): strain variation in the nonstructural gene region en
A:Reference number: A48547; MUID:92271462; PMID:1589964
A:Accession: B48547
A:Molecule type: genomic RNA
A:Residues: 965-1691 <FRY>
A>Note: sequence extracted from NCBI backbone (NCBI:104576, NCBI:104578)
C:Superfamily: hepatitis B virus nonstructural protein
C:Keywords: ATP; GTP binding; nonstructural protein; nucleotide binding; nucleotidyltr
P:973-980/Region: nucleotide-binding motif A (P-loop)
P:979/Binding site: ATP/GTP (Lys) #status predicted

Query Match	6.1%	Score 542;	DB 1;	Length 1691;
Best Local Similarity	21.8%	Pred. No. 9.6e-24;		
Matches 382;	Conservative 186;	Mismatches 554;	Indels 632;	Gaps 78;
2 YAKATDVAAVRYAAADVAYA...LQRAVYKLP...LAKGGLPRT 59				
6 FIKAGCITTAIQALAAANALANAVVRPLSHQVILINLQPKQLVFRPVP--W 63				
60 OHPIIAGHQAEEVYAHNPARGRS--TVLEIGPSLHSAUKGAPNAPVADYGCCTKYG- 117				
64 NHPI-----QRVHNLBQYCRASGRCEIGHNPS---INDNPNV-----LHRCFLHPVG 112				
118 RQGSRHITNLR-----SRVATGRPEPKADSLANGIARTECVDCVGSAPKSRVG 170				
113 RDVQRMVYAPTRGPANCRSRSLRGIP-----ADRTYCFDGFACCPAAETG 160				
171 IANHSIVDTLEELANAFENHGLAHVRAFMHPEBELLYMDNVNNAELGTRFVIBPMNV 230				
161 VALYSLHQLPADVAEMARHGMTRLVLAFLHPEVLL-----PGGY 203				
231 KDAFOGQDGLHPELDFINESQRRIRERLAARGSYSRAVIFSGDDWDGAYLHD--- 287				
204 RTSSY-----LLIH-----DGKRAVVTYGGDTSAG--YHNDVAT 235				

```

QY 288 FHTWL-AYLVNRYPTPEGFSLHIEVQRHSSIELRTRAP-----PGRMLAVVP 338
DB 236 LRTWTRTKVGEHP-----LVIERVRIGCHFVLLITAAPEPSMPVVPYPRSTEVV 289
QY 339 RTSQGLCRIPNIFYVADASGTEHKTILTQCHKNMMLNFMQTRPEKELVDMVLASFARA 398
DB 290 RSIFRGSGPSLFPFACAVKSTFTHAVPT--HIMDRMLMFGALTDQAPC-CRLMTYLRG 346
QY 399 RLRAIVASEVTSSWNISPADLVRTVSLY--VLH-----I 433
DB 347 ISYKTVGALVANEGNATEDALTAVITAYLTIQHRYLTQATISKGRRLLEHAQKF 406
QY 434 IERRAAVAVKAKDV-----RGSTFMS-----LKV 463
DB 407 ISRLSWLFEKSGRDIPEGROLQFYAQCRMLSAGFHLDPRTLVPDESVCSCRTTIRI 466
QY 464 LGS-CCGLENL-----KGTDV-----VFTRVVD-- 486
DB 467 AGKFCCEMKLGOEBCSFLQAPAGLAGDGDHNEAVESDVTAPATIDITGSYIVDGR 526
QY 487 -KRYVHSLGDIICDV--RLS----- 503
DB 527 SLQTYQALDLPADLVARAARLSATVTVTETSGRLDCQTMIGNKTEFLTTFVDGARLEVNG 586
QY 504 PRQVGF-----LPSRV-----PPARYP 520
DB 587 PEQLMLSPDSQOCMAAGPFLTYAVDGLVHVFSTAGLSRVVPPGNAPTAPSEVT 646
QY 521 -----HDRE----- 525
DB 647 AFCSALYRNROSQSVIGSLMHEGLGLFPFSPGHEKRSANPFGCESTLYTRTWS 706
QY 526 -----LEVLRAGCYNERPVSTPV----- 546
DB 707 TITDPLTVGLISGHLDAAHPSGSPATATGAVGSDSDPDPPLDVTGDSRPGARA 766
QY 547 -----BEPQG 551
DB 767 GNPNGVPOQRLLHTYPDGAKIYVGSIFSECTMVLVNASNAGHPGGGLCHAFQRYPDS 826
QY 552 FDA-----DLNHAATASLPEYR-----ATLOAG----- 574
DB 827 PDATKFWMDGLAAYTLTPRPIIHANA---PDRLEHNKRLKLEAIRETCARGTAAYEL 883
QY 575 LNTDVQKQKLT-----ENALKTIDGLTSPV----- 602
DB 884 LAGIYQVPSLSFDMERNHRRPDELTYLTELAAWFESNRPGOPTLNTEDTARAANDA 943
QY 603 -----GLEMYE-----GPRSGKTGLIALBAAGKALVYATREL 639
DB 944 LEIDSGSEVGRACAGKVEPVGVVRYOFTAGVPSGSKSVQA-----DVDVVVPTREL 998
QY 640 REAMDRIRIPASATQHYALATL--RRATAGAPATVYIDCFMFPYVYVALV 697
DB 999 RNAMRRR---GRAAFPTTHAARTSGR-----VIDABSLP--PHILLH-MQR 1043
QY 698 SSRIVLVGDVHQIGFIDQGSAMPVLVDVVKQCRRTFNOQTRCPADVATTPFQSIY 757
DB 1044 AASVHLGDPNQIPALDPEHTGL-IPAIR--ELVFTSMWHVTHRCPADV--CELVRGAY 1098
QY 758 PGCTTSGCVASISHVAPDRNSAOQLCTQTEBKRHGAEGAMTYHEAQRFPAVILH 817
DB 1099 PKIQTTSKYLRSLFMBGP---AVGQKLVFTQAANKAH--PGSITHEAGATFTTTTII 1152
QY 818 YNGSTAEOKLAEKS--HLLVGTTRHTNHLXIRDPTGDIEROLNHS-----AKAEV 866
DB 1153 ---ATDARGLISSAAHIVALTRTBKCVILDSRG-LRGVIGIDALVNNFSLSGEV 1208
QY 867 FTDIPAPLEITVYKPSBEVORN--EWATIPPOGATPHGAILHLKRNFGQPDGCVYALAK 925
DB 1209 GHQRP-----SVIPRGNPDRANDVLAAPPSQCI--SAFHQABEIGHRP--APVAALVP 1259
QY 926 TGEVVGRAKINVELAEPAATPKPHRAP--EGVQVNTNMSNGIALQTLISRTYKS 984

```

```

DB 1260 PCEBLGQILYIPGLASCDVSV-----TFBLTDYHCSMAABSORKAVLSTLVGRKRT 1315
QY 985 A--DLPLHAKEDVYKMLNSLDHNMWTVTEBDARDAVFEQLKFTORGTVEDLLEPD- 1041
DB 1316 RLYDAGHTDVASLARPIFLTOR-----VT--ATCGELPELVBAVNBKQDQSAVLELDL 1368
QY 1042 -DPYTRDIDFLMKTOOK-VSPRPINTGVGGGIAHSHSLNVLAAWRILES-ILRTGS 1098
DB 1369 CSRDSRRTTFPGKCNKKTGTGTJAHGVGQOI FWSKTFCLFGPWFAIKALISLSP 1428
QY 1099 RTVRSNGLPDEBEAMLKAKINQVHATPVASADTBEPDTAHNNTSLLFALLBRIGTP 1158
DB 1423 QAVFYGDADYDS-----VPSAAYAGASHAMVENDSEBDSQNNSTSLGJLMECCMP 1484
QY 1159 AAANVLFRR-----RCQKRTLRA---KGLGSEVVDGLDLSGAAMPTRNTIFSAANVL 1208
DB 1485 QWLVRLYHAVNSGAWILQAPKBSLRGFMKKHSG--BPSGL-----W-----NTVMMAITA 1533
QY 1209 TL--FRGVKPAAFKQDGS-LICGSHYLRPDSRLMGBRYTKHLKVEYQKIVPYIGLLV 1265
DB 1534 HCYEPRDIQVAAPFKGDBSVYLCSEYRQSPGASLIAGGLK---LKADFRPIGLYAGVV 1590
QY 1266 SABQVVL-DPVRSA 1278
DB 1591 APGLALPDVVARPA 1604

RESULT 2
NNNNNN
genome polypotein - hepatitis B virus (strain Burma)
N/contains: RNA-directed RNA polymerase (BC 2.7.7.48)
C/species: hepatitis B virus
C/date: 30-Sep-1992 #sequence__revision 30-Sep-1992 #text_change 23-Jul-1999
C/accession: A40778; A48547
R/tam: A.W.; Smith, M.M.; Guerra, M.E.; Huang, C.C.; Bradley, D.W.; Fry, K.E.; Reyes,
Virology 185, 120-131, 1991
A/tile: Hepatitis B virus (HBV): molecular cloning and sequencing of the full-length
A/reference number: A40778; MUID:92024067; PMID:1926770
A/accession: A40778
A/molecule type: genomic RNA
A/residues: 1-1693 <TAM>
A/cross-references: GB:M73218; NID:9330023; PID:AAA45734.1; PID:9330024
R/fry: K.E.; Tam, A.W.; Smith, M.M.; Kim, J.P.; Luk, K.C.; Young, L.M.; Platak, M.; Fe
Virus Genes 6, 173-185, 1992
A/tile: Hepatitis B virus (HBV): strain variation in the nonstructural gene region en
A/reference number: A48547; MUID:92271462; PMID:1589964
A/accession: A48547
A/molecule type: genomic RNA
A/residues: 967-1693 <FRT>
A/cross-references: GB:M32400; NID:9330021; PID:AAA03206.1; PID:9330022
A/Note: sequence extracted from NCBI backbone (NCBIN:104572; NCBI:104573)
C/superfamily: hepatitis B virus nonstructural protein
C/keyword: ATP; nonstructural protein; nucleotidyltransferase

Query Match 5.4%; Score 479; DB 1; Length 1693;
Best Local Similarity 21.2%; Pred. No. 5.6e-20;
Matches 369; Conservative 188; Mismatches 577; Indels 610; Gaps 73;

QY 2 YAKATDVAVAAADVAVVNVVQRAVYKDFAPPLKALTELRLYVP--LRFPGTLPPT 59
DB 6 FTKAPGITTATIGALAAANNSALANVVRPLSHQOILTNMQPRLVRRPVP--W 63
QY 60 QHPITAGORVAEVLNPARGRS-TVLEIGSLHSALTKAGAPNAPVADYHGC-TKYGT 117
DB 64 NHPI-----QRVHNEBELYCRASGRCLTGAHPS--INNPNV-----VHRCFLRPVG 112
QY 118 RQGSRHITLLE-----SRVATGRPERKADASLIANGIASRTFVGVGSCAFPSRVG 170
DB 113 RVQQRVYTPTRGPANCRSALRGLP-----AADRITYCLDPSGQNPFAEYTG 160
QY 171 IANHSLYDTLELANAFENHGLFMVVARFMRHPEELLVMDNVVNAELGYRFVYIEPMAY 230

```

```

Db      161 IALVSLHDMSPDVAAEMFRHGMTRLYALHLPREVLL-----PGCTY 203
QY      231 KDCAFQGGDLRLHPELDFINESQERRIERLAAGSYRRAVIFSGDDMDGDAYL---- 287
Db      204 RTASY-----LH--DGRV-----VVTYEGDTSAG--YNHDSN 235
QY      288 FHTMAYLVNVPPTFGFSLHIEVQRHSGSIBLRTRAP-----PGDRLMAYVPR 339
Db      236 LRSM-----IKTKVYTGDPHVIIEVRVAIGCHFVLLLRABPSPMPVPPRSTREYVR 290
QY      340 TSQGLCRIPNI FYVADASGTEHKITLTSQHKVNMMLNFMQTRPEKELVDMTVMASFAPAR 399
Db      291 SIFPGGTPLF--PTSCSTKSTFHAVPAHIMDRMLFGATLDDQAF--CSRLMTYLGII 347
QY      400 LRAIVASEVYESNNISPADLVTRVSLV-----LHIERRAANAV 443
Db      348 SYKVTGTVLWNEGNNASEDALTAIVTAAYLTICHQRYLTQALSKGRRLREHQAQFI 407
QY      444 -----KTKADVPF--ETSFMESLKHVLS----- 466
Db      408 TRLSKMLFEKSGRDYIPGRQLEFYAQCRRWLSAGHLDPRVLVPDESAPCHCTAIRKAL 467
QY      467 -----CC-----GLNNLKGTDVVTKR 483
Db      468 SKPCCFMKMLGOECTFLQPAEAGAVGQGHNEAVESDVPDPAESALSDISGSYVVPETA 527
QY      484 VVDKRYHSL-----GDIICDVR-----SP 504
Db      528 LQPLVQALDLPAEIVARAGRLTATVKVSQVDGRIDCETLGNKTFRTSPVDGAVLETNGP 587
QY      505 EQ-----VGFLPSRVP----- 515
Db      588 ERHMLSPDASQSTMAAGPFLTYAASAAGLEVRYVAAGLDHRAPAPVSPRSRAGEVTA 647
QY      516 -----PARVPH----- 522
Db      648 FCSALYRENRBAQRHSLIGNLHMFHEGLIGLFAFPSGHVMSANPCGSESTLYRTWSR 707
QY      523 -----RELEVLREAG-----CNERPVSTPVEBPOG----- 551
Db      708 VDAVSSPARPDIGFMSSEPSIPSRAATPTLAAPLPPADPSPPSAPALABASGATAGA 767
QY      552 -----PDADLMHAT----- 560
Db      768 PAITHQTKARRRLFTYDGSKVFAAGSLPESTCTLVNANVDNRPGGGLCHAFQRYPA 827
QY      561 -----AASL-----PEYR-----ATLQAG-----LN 576
Db      828 SFDAASFVMDGAAYTLTPRPIHVAVADYRLBNPRLEAAVRETCSRIGTAYPFLG 887
QY      577 TDVQO-----LKIT 585
Db      888 TGIQVPIGSPFDAMERNHRPGDELYLPELAARWFEANRPTPTLTTEVDARTANLAIIE 947
QY      586 LENL-----KTIDGLTSLPVGLEWYE---GPGSGKTGLTALAAGKALYVAPTEL 639
Db      948 LDSATDVGACACGCRVTP--GVVOYQFTAGVPGSKSISITOA-----DVDVVPFTEL 1000
QY      640 REAMDRRIKPPSASATQHVALAIL--RRATAGAFATVVIDECFMPLVVAIVHALSP 697
Db      1001 RNAMRRR---GFAATPHTAARVTOGR-----VIDEAPSLP--PHILLH--NQR 1045
QY      698 SSRVLVGVADVHIOIGTIDQGSANMPLVVDVVKQCRRTFNQTKGCPADVATTFFQSLY 757
Db      1046 AATYHLLEDPNQIIPAIIDEHAGL--VFAIRPDIGP--TSMWVYTHMPADV--CELRIGAY 1100
QY      758 FCGTTSGCVASISHVADPYNNSQOQTLCTQOEBSRBAAGAMVHAAGRTFASVLH 817
Db      1101 PMIOGTSVAVLSLFWGER---AVGQKLVFTQAAKPAK--PGSVVHAAGATTYETITII 1154
QY      818 YNGSTAEOCKLAEEKS--HLVGTIRHTNHLVIRDPGDIERQLNHSAAAEVFTDIPALE 875
Db      1155 ---ATADARGLIQSSRAHAIVALTHTTEKCVIIDAAPG--TLREVGIS--DAIVNPFLLAGE 1209

```

```

QY      876 ITTVKPSSEVQRN-----EVMATIPQSATPBGAIHLIRKNRPGDPDGCVALAKTGEV 930
Db      1210 IGHQRPSTVIRPGANDAVDTLAAPFPGQI--SAFHQLABELGRP--VPVAIVLEPPCEL 1266
QY      931 FGRBAKINVELAEPDAPPKPRAPQ--BGQWVKVYTNASNKHQALQTLISRYTKSADLP 989
Db      1267 EQGLLYIPQBLTQDSV-----TFELTDIVACMAAPBQRAVASTLVGRGRT---KL 1319
QY      990 HEAK--EDYKRLNLSLDHMTWTTEADRAVFEETOLKFTORGCTVEDLEPD--DPYIR 1046
Db      1320 YNASHSDVR---DBLAFIPAIIGVOVTTCELLEYELVEAMVKGDGSAVLELDLCNRDVS 1376
QY      1047 DIDPLMKTOCK--VSKPRTNGKVGQGIASHGSLNPLVLAARILIE--ILNTRGRTVYS 1104
Db      1377 RITFPQDCNKFPTTGETIHNQKVGQGISAMSKTICALGPPFRIETKILMLPQGVPG 1436
QY      1105 NGLPDER--EAMLEBAKINQVPHATFVSADMTVEPTAHNNTSBLPAAILERRIGTPAAVN 1163
Db      1437 DAFDDTVSAVAABAASMW-----FRNDFSEPDSTQNNFSLGBCAIMECGMPOMLIR 1491
QY      1164 LPRERCGKRTIPANGLSVEVDGL-----LDSGAAMPCCNRTIFSAAVMTL--FRGVK 1216
Db      1492 LY-----HLIRSAWILQAPKESLRGFWKSGSGPGLTMMNTVMMAVITCYDPRDQV 1545
QY      1217 AAFKGDSTL--ICGSHYLRFDASRLMGERYTKIKLYEVOKIVYIGLVSABQVYL--DP 1274
Db      1546 AAFKGDSTVLCSHIRQSPGAVALIAGGCLK--LKQDFRPIGLYAGVAVPGLGALPDV 1602
QY      1275 VRSA 1278
Db      1603 VRFA 1606

```

RESULT 3

```

NM17681
180K protein - tomato mosaic virus (strain I)
N:Contains: 130K protein
C:Species: tomato mosaic virus
C:Date: 28-Aug-1985 #sequence_revision 28-Aug-1985 #text_change 30-Sep-1993
C:Accession: A04195
R:Ohno, T.; Aoyagi, M.; Yamanehi, Y.; Saito, H.; Ikawa, S.; Meshi, T.; Okada, Y.
J. Biochem. 96, 1915-1923, 1984
A:Title: Nucleotide sequence of the tobacco mosaic virus (tomato strain) genome and c
A:Reference numbers: A91984; M01D:85157522; PMID:6549393
A:Accession: A04195
A:Molecule type: genomic RNA
A:Residues: 1-1615 <OHN>
A>Note: readthrough of the terminator UAG between codons for Gln-1116 and Gln-1117 occ
C:Superfamily: cucumber mosaic virus RNA 1 protein

```

```

Query Match      3.7%; Score 323; DB 1; Length 1615;
Best Local Similarity 19.6%; Pred. No. 1.1e-10;
Matches 316; Conservative 193; Mismatches 520; Indels 586; Gaps 72;

```

```

QY      24 QORAVKIDRAPPLALETL--HRLYPLRFKGTLPFPGHRI--LAGQRYAE----- 72
Db      42 RDRRPKNVFSKVSVEBQTLATKAYPEQI--TFYNTQNVHSLAGLSLELEYLMMQ 98
QY      73 -----EVLHNP-----RGRSTVLAIGPSLHSLAKLHGAAPNAVADYHGTCTKYGTBD- 119
Db      99 IPYGSITVDIGGNFASHLFRKRAY-----KCM-----PNLDVRI--MRHEGQKDS 144
QY      120 -----GSRHITALESSVATGRDEFPADASLLANGIASRTFCVGVGSC-----A 164
Db      145 IELYLSRLERGNKIV-----PNFQKAFDRYABKPNVAVCHDTFQTCNHSQEC 192
QY      165 FKSRV--GIANSKYDVTLIELANA FENHGLMVAFAFMMBELLYMDVVNVAALGRTYHV 223
Db      193 YTGRTVALHSHSTIDIPADEGALLKRNVAVCTAFAFPSSRLLEDSHVIND----- 245
QY      224 IEBPNAVDCAFQGGDLRLHPELDFINESQERRIERLAAGSYR-----RAVIFSGDD 278

```

Db 246 -----EINACFORGDG-RLTF-----SPASES-----TLNYSHSYNIILKVCCTYTPASN 290
 Qy 279 DWGDAYLHDF-----HTWL-----AYLLVR-----NYPTPFGE 306
 Db 291 --REVVMKEFLVTRVNTWFCFGRIDTFLLYKGVAAKGVDSBQFYKAMEDAMHYKKTLLAM 348
 Qy 307 SLHIEVQRHSGSIELIRTPARPGDBMLAVNPRTSGQLCRIPNIFYAASGTEHNTILT 366
 Db 349 CINSERLLEDSSSVNYP-----PKMDMVLPFLDISL-----ETSKTRREVLV 394
 Qy 367 SOHVMMLNFMQTREREKELVDMTVLSFAPARLRAIVASEVTESSWNISPADLVRTVV 426
 Db 395 SDOFYTVLNHRTYQAKLTYSNVL-SFVESIRSKVINGVYARABEMVDKS-LQOSLS 452
 Qy 427 SLVYLHI-----IERRAANAIVKTAADVFGETS--FWESLKVVLSCGGLNMLKGT 476
 Db 453 MPEFHTKTLAVLKDDLLISKFALGPQVSGHVDEISLAFGNAPSIKERLIRKLKIKIT 512
 Qy 477 D-----VFTKRVVDKRVHSLGDIICDVNLSPEQVGLFSRVPAPVPHDRER 525
 Db 513 ENALEIRVDPDLYTFPHDLVSEYKM-SVDMPLVDIRKMEET-----EEMYNALSE 562
 Qy 526 LEVLRAGCYN-----ERP----- 539
 Db 563 LSVLKNKDKRVDVFGQMOCSLEVDPTAKAVIVAVMSNESGLTTFEQPTANVALAQ 622
 Qy 540 -----VPSTPVEEPQ-----GPDADI-----WH-A 559
 Db 623 DSEKASDGLAVTSPRDEEPSIKSGMARGELQLAGSDPESSYTRSEIESELQFHHMA 682
 Qy 560 TAASLEBYR-----ATLOAGLNTDVKKQK- 583
 Db 683 TASSLHKQKCSIVYTGPKLVQOMKNEFIDSLVSLAASVNLKILDTAIDLRTROKF 742
 Qy 584 -----ITLEN----- 588
 Db 743 GVLVDVASKRWLVKPSAKNHAQVETHTARKYHALLHEDFGIITCDNMGRVAVSSSVY 802
 Qy 589 -----ALKIT-----DGLTSPVRGLEMYEGPQSGKTGTLAALBAAGKALYVATRE 638
 Db 803 YSDMAKRLTLRLKGEPRHVSAAKAVLVGVPCGCKTEILSRVNE--EDLILVGRQ 860
 Qy 639 LREAMDR-----IKPESASATQHVALAIIIRBATABAPATVVIDCEMFPLVYALIV 693
 Db 861 AAMIRRRANASGIVATKDNVTRVDSFLNNYGGKARCQFKRIFIDGLMLHGTGCVAFV 920
 Qy 694 ALSPSRIVLVGVHVGIFID-----FOGTSANMPLVADVVKOCRRRTFNOGTCRPAADV 749
 Db 921 EMELCDIAVYVGTQOIPYINRYTGFPYPAHFAKLEBDEVE--TRRT--TLRCPADV-- 973
 Qy 750 TTFPOSILYPC-CTT-----TSGCVASISHVAPDRNSQAOTLCFTQSEKS--RHGA 797
 Db 974 THFLNORYEGHWACTSESEKSVSQEMVGAASINPVSKPLKGIITFTQDKEALLSRGY 1033
 Qy 798 EGMATVHEAGRTFPASTIILHNGSTAQKLAESHTLNGITRHTNL-----YLRD-- 850
 Db 1034 ADVHTAHEVQGETYADVSLVRLPTPVSIIRADSPhVLVSIRTSKLTYYTVVMDPLVS 1093
 Qy 851 -----TGDI-EROL----- 858
 Db 1094 IINDLERVSSYLDMYKVDAGTQQLQVDSVPKPNFLVAPAKPGDISDMGFYDCKLPGN 1153
 Qy 859 -----NHSAAVFTDI-----PAYLEITTVKSEEVORNEVWATIPQ 898
 Db 1154 STLLNNYDAVTMLKTLISLNVKDCIILMSKSVAPAKD--VKPT-LIPMRTAAEMBRQT 1209
 Qy 899 ATRHGAHILKRNKGGDPDQGCVALAKTYGVFVGRKAKINE-----LAEPRATPKH 951
 Db 1210 GLEENLVAMIKRNNSPELSGVNDIENT--ASLVVDKFPDSYLLKE--KKEPN 1258
 Qy 952 RAFOEGVQWVTVASNKHQAOLQTLRSYTKRSADLPLHEAKEDVDKMLNSLDHMDMTV 1011
 Db 1259 KNF-----SLFSR-----ESLNR--MIA 1274

Qy 1012 TEDADRAVRFQQLKFTQRGGVTEBDBEDPYRIDIDPLMKTQ--QKVSPPKINTGKV 1069
 Db 1275 KQE-----QVIT-----GQLADFPDVLDAVVOYRRMITAQAPOKQKLLSIQTEYPAL 1321
 Qy 1070 QGIAAHSKSLNFMVLAAMIRIIEILR-----TGSRTVYNSGLPDEBEMMLBEAKINGV 1123
 Db 1322 QTVVHSKTKINAFIP--LPSELTRQLDSDSRFLPFTKTPAQIDPFQDLD-- 1374
 Qy 1124 PHATVPSADWTFPTAHNNTSILPALLR-----RIGTPAAVNLFRERCGRTILRAGK 1178
 Db 1375 ---SHVPMOVLHLDVSKYDKSGNBFCFAVEYIWRILGLJEDFLAIVWKQGHKTKTKDYT 1431
 Qy 1179 LGSVEYDGL-----DSGAAMTPCKRTIPSA---AVMLTLFPGVKRAAFGDDSL 1226
 Db 1432 AG---IKTCLWYORKSGDVTYTGIVTIASCLASLPMKLIK-GAFQGDSDL 1482

RESULT 4
 genome polypeptin - Chinese rape mosaic virus
 N/Alternate names: RNA replicase
 N/Contains: RNA-directed RNA polymerase (BC 2.7.7.48)
 C/Species: Chinese rape mosaic virus
 C/Date: 19-Mar-1997 #sequence_revision 29-Aug-1997 #text_change 17-Mar-1999
 C/Accession: S65053
 R/Author: I. Sanchez, F. Martin Martin, A. Martinez-Herrera, D. Ponz, F. Plant Mol. Biol. 30, 191-197, 1996
 A/Title: Nucleotide sequence of Chinese rape mosaic virus (oligoseed rape mosaic virus),
 A/Reference number: S65053; NCID:96197410; PMID:8616237
 A/Accession: S65053
 A/Status: nucleic acid sequence not shown; translation not shown
 A/Molecule type: genomic RNA
 A/Residues: 1-1597 <AGU>
 A/Cross-references: EMBL:U30944
 A/Note: readthrough of the terminator UGA occurs between codons CAA for 1103-Gln and C
 A/Note: the internal stop codon is translated as X
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, July 1995
 C/Keywords: nucleotidyltransferase

Query Match 3.4%; Score 298.5; DB 2; Length 1597;
 Best Local Similarity 19.0%; Pred. No. 3.1e-09;
 Matches 287; Conservative 184; Mismatches 524; Indels 515; Gaps 60;

Qy 171 IANSLDYVTLBELANAFENHGLMVRAPMPEER-LLYMDVNVNAELGYRPHVIBPMA 229
 Db 199 IALHSYDIPADFGAALLRKVKYKICYAAFHFIENMLDQDSVTLEDIGATFO----- 251
 Qy 230 VMDACQGGDLRLHPELDLPINESQRRIRBLAARSGSRRAVTPSGDDDKGDAYLHDPH 289
 Db 252 -----RAQD-KLNF--PFHNESTLN-----YTHSFS 274
 Qy 290 TWLAYLIVRVPPFPFSLHIEVQRHSGSIELIRTPARPGDBMLAVNPRTSGQLCRIPN 349
 Db 275 NIIRKYCKTFPFAKQYVYHKEP-----LVTRVNTWYCKPTR 311
 Qy 350 I-----PYTA----- 354
 Db 312 VDFTLFPGVYKTSVDSSEFYKAMDADWEYKKTILAMLSERTIFKDSAAINFWPKVRDM 371
 Qy 355 -----DASGT-----EHKITLTSQHVMNLNFMQTREREKELVDMTVLSFAPARLRAIV 404
 Db 372 VIILPFDASTTGRMSRRRLVANKDQVYVTLNHIKTYQAALTYANVL-SFVESINSRVI 430
 Qy 405 VASEVTESSWNISPADLVRTVVSLVYLHIERRAANAIVK-----TADDFGETSFV 457
 Db 431 INGVTAASBMDTKALILGPLAMTFPIVTLGSHQDSDIVLKKPKQPRATAKELIW--SSLC 488
 Qy 458 ESLKHYLGSC-----CGLNMLKGTDV-----VFTKRVVDKTR- 489
 Db 489 DALKGIVPSVGETILARGGFVQLAEBSLRIKIPELYCTFTDRVLVLYKRTREBFSQCDLSKP 548

```

QY 490 -----VHSLGDIICDV-----RLSPQGVG-----FLPSRV 514
Db 549 LEESEKYNALSELSESVLENLGSFDLDKFKELCQKKKNVDPVAAKVVAIMNSSELTLPKK 608
QY 515 PPAVFDREBELEVLREAG--CYNERPVSTPVEE-----POG-----FDADL 556
Db 609 PTEEEVABALSGEVQDEGLRLSNKAPPPCVSNLEGLVPAQGLCPNGANFDRVMDISE 668
QY 557 MHATAA-----SLPEYRATLQAGLNDYK-----580
Db 669 FHLKSVDAAVKKAMMSAVYTGKIKYQKKKNVYDYLASLSTVSNLCKRYLDVHGVDPS 728
QY 581 -----OLKITLEN-----588
Db 729 QEKSGVMDVRRGRMLLNKNAKCHANGVAEDANHKIVYLLMWDESKPYCDETFWFLANSS 788
QY 589 -----ALKTI-----DGLTSPVRGLEMYEGEPGSGKTGTLIAALEAAGKALYYA 634
Db 789 DSLVYDWGKLTILTSCCRDGEPPFPKLVLVVDGVPCCGKTKELKYNFS--EDLVLY 846
QY 635 PTEIREAMDRRIKPPSASATQVHALILRRATAGA-----PPATVVIDEC 681
Db 847 PGKASAKMIIR-----ANQ-----AGITRADKNDVTVDSFLMHPKRVFKRLFIDEG 895
QY 682 FMFPLVVAIVHALSPSSRIVLVGDVHQIGFIDFGQTSANMP-----LVRDVYKQGR 734
Db 896 LMLHTGCNVFLMLSHCDVAVYVDTQOIPFI--CRVANFPYPAHFAKLVVD--EKEERR 951
QY 735 RTFNQKRCPADVAVATPFPQSLYPG--CTT-----TSGCVASISHVAPDYRNSOATL 785
Db 952 VTL-----RCPADV--TYFLNQKYDSVLCSTSVSVAEVRGKALNPIITLPLEGKITL 1005
QY 786 CFTQGEKSR--HGAEGMTYHEAOGRTFASVYILHNGSTABEQKLAESHLVIGTHT 842
Db 1006 TFTQADKFKELLDKGYKDVNTYHEVGSEYKETAIRLATPLEIISRASPHVLVALTHTT 1065
QY 843 N-----HLVIRDP-----TGDIERQLNHSKAEVFTIIPAPLE-----ITTVPS 882
Db 1066 TRCKYTYTVLDPMNANVISLGLSNFLBMTKVESGTOXQOQIDIVFKGTMLFVTPKSG 1125
QY 883 EEOGNEVMATIPPOSATPHGAIHLIRKQFGD--OPDCGVALAKTGYEFGGRAKIN 938
Db 1126 DMRDQFVYDVLTPENSTILNEFDVMTNLRDISLNVDC-----R 1166
QY 939 VELAEPDTPKPHRAFOGVQVKTNAENKQALQTL--LSRYTKRSADLPLEHAKEDV 996
Db 1167 IDFSKSVQVPERPVFMK--PKLRTAAEMPRTAGLENLVAMIKRNNAAPDLGTGIDI 1222
QY 997 KRMLN--SLDRHMDWTVTED-----ARDRAVFEQOLKFTQRGGVEDLLR--PDDPYI 1045
Db 1223 EDTASLVVEKTFMDAVVKEPSTGDMAMTRSFSSWLS--KQSSSTVGQLADNFVLDPAV 1281
QY 1046 RDIDFLMKTQ--QKVSPPRINTGKVGOGIAHSKSLNFIYLAAMI--RIIEELIRTSR 1099
Db 1282 DEYKMTMSQPKQKDLDSIQDEYRALQITI VHSSKIKINAFGPMSELRMLLETIDT--SK 1340
QY 1100 TVRYSNGLPDBEBAAMLBAKINQVPHATFVSADWTEPFTAHNNTSELLFAALLERIGPA 1159
Db 1341 FLFYTRKTPTQIEEFPSDLDSQA--MEILLEDISKYDSQNEFCAVEYKIMKLGIDD 1398
QY 1160 AAVNLFRRCCKRTIRAKGLSGVEVDGLL--DEGAAMTPCRANTIPSAAV--MLTLFR 1212
Db 1399 WLAETWRCQHRRTTKIDYTAG--TKTCIMYORKSGDVTTFIGNITIIIAACLSMIPMDK 1455
QY 1213 GVKFAFAKGGDSLL-----CGSHYLRDASRL 1239
Db 1456 VIK-AAFGCDLSLIYIPKGLDLPDIOAGANLTMNBEAKLFRKKYGVFCGRYIYHHDRAKI 1514
QY 1240 HNGERYTKTKHLKNEVQKIVPIYGLLSAEOVVLDVRSALKIFGRC-YTSEL--LYSK 1294
Db 1515 -----VYDPLKILSKGCKHIRDEVLIELRSL-----CDVTSNINCAVYSQ 1559
QY 1295 YVEAVRDIITK 1304

```

```

Db 1560 LDEAVAEVHK 1569
RESULT 5
MMTMFV
183k protein - pepper mild mottle virus (strain Spain)
N:contains: 126k protein
C:Species: pepper mild mottle virus
C:Date: 30-Sep-1992 #sequence_revision 30-Sep-1992 #text_change 08-Apr-1994
C:Accession: J01312
R:Alonso, E.; Garcia-Luque, I.; de la Cruz, A.; Wicke, B.; Avila-Rincon, M.J.; Serra, J.
J. Gen. Virol. 72, 2875-2884, 1991
A:title: Nucleotide sequence of the genomic RNA of pepper mild mottle virus, a resistar
A:Reference number: J01312, MUID:92115526, PMID:1765765
A:Accession: J01312
A:Molecule type: genomic RNA
A:Residues: 1-1611 <ALO>
A:Cross-references: GB:M01413
A>Note: readthrough of the terminator UAG occurs between the codons CAA for 1117-Gln at
C:Comment: This protein may have RNA polymerase activity.
C:Superfamily: cucumber mosaic virus RNA 1 protein
F:1-1117/Product: 126k protein #statue predicted <PRO>

Query Match 3.4%; Score 298; DB 1; Length 1611;
Best Local Similarity 19.6%; Pred. No. 3.4e-09;
Matches 318; Conservative 188; Mismatches 542; Indels 574; Gaps 77;

QY 10 RVYAADAVYANVLCQRAVKLDPAAPKALSTL--HRLVYPLRFKGTLPPTQHPI--LA 65
Db 30 RLYESA--VEQNA--HRRPKVNFARSISEGTLATAYPERQI--TFYNTQNAVHSLA 84
QY 66 GQGVAR--EVLNARAGSTYLETGPSLHSLKLGAPNAVADY-HGC-TKYGTRGDSR 122
Db 85 GGLSLKELEYLMOQIPYSTIYDIGNP-AAHMERG-----RDYVHCMMEMDRLRDYMR 137
QY 123 HITLBSRSVATG-----REFPADSLANGIASRPF--CVDGVSQAPKRS 167
Db 138 HNAQDSIELYLSKLAQKKVYIPYQKCFPKYTTDDPOSVCSKPFQHC--GVSHTCDKV 196
QY 168 RVGIANSLYDVTLEELANAFENHGLHVNARFEMPEBELLYMDNVVNA--ELGYRHYIEE 226
Db 197 -YAVASHLYDIPADEPFAALRRNVHCVAAFPHSBELLEDVSYSLDIDGAPFS-----251
QY 227 PMAVKDCAFQGGDLRLHPELIDFINESQGERIERLAARGSYSR-----RAVIFSGDDWG 281
Db 252 -----REGDM-LNF--SFAVES-----TLNTHSYSNVLKYCYKTFPAGS--R 290
QY 282 DAYLHDF-----HMTL-----AYLVLR-----NYPTPPGSLH 309
Db 291 EYMKERFLVTVNVTWPFCKPSRLDTFVLRYGVYHGVDRKQFYASAMEDAMHKKTLAMONS 350
QY 310 IEVORRHSSIELNITRAPPGDRMLAVPRTSGCLRIPIFYVADAGTEHTILTSQH 369
Db 351 BRILLESSSYNMYF-----PKMKQNVYF-----LPDVSLONEGRLARKETMVSQD 398
QY 370 KVNMLNFMQTRPEKELVDMTVLMSFAPARLARIVASVTESSWNISPADLVRTVSLY 429
Db 399 FVYTVLNIHRTYQSALTYANVL--SFVESHSRVLYINGVTARSEMDVDKA-----LLQSLS 453
QY 430 VLIHIERRAVAVKTAODVFGF-----TSF--WBSLKHVLGSCC-----468
Db 454 MTFPLQTLANL-----KDLVVOQFOVHKSLEYVDKETAFAFNCFPPIKERLINXK 508
QY 469 -----GLNLIKGTD--VFTKRVVDKYRVSGLDIIQDVRLSPQGVGFLPBRVPARVF 520
Db 509 LITVSEKALKEIKVPLVYTFHRLVLYKEIK--SVEKMPVLDVYKSLSEABAVMNALEISIL 567
QY 521 HDREBEL-----VLRAG--CYNERPVS-----542
Db 568 KDSKFDVDVFSRMCNTLGVPLVAKVMAVVSNSGLTTLFERPTEANVALAQPTIT 627
QY 543 -----TPVEEPQGF-----553

```

```

Db      628 SKEGSLKIVSDVGESSIKEVVRKSEISMGLTGVTSDEFORSTEISLQGFHNVSTE 687
Qy      554 ---ADLMHATASLP-----EYRATLOAGINTVQKOL-----ITLEN----- 588
Db      688 TIIRKQMAHMYVTGLPKVQCCKNYDLSLVAISAAVSNLKKIIKOTALDLSTKEKFGY 747
Qy      589 -----ALKTIDG-----LT 597
Db      748 DVCLKKMLVKPLSKGHAMGVMDSDYKCFVALLITYGENIVCGETWRVAVSESLVSD 807
Qy      599 LSPVGLG-----MYEGPBGSKCTLTALAEAGKALVYAPRELA 641
Db      808 MOKIRAIRSLVDGEPIHSAKVTLVDGVPGCCGKTEILSRVNF--DEDLVLVPGKQAA 865
Qy      642 AMDRRIKPPSASATQVALAILRATAGAPAT-----VVI 678
Db      866 -----MIRRRANSGLIYATKENVATVDSFLMNYGRGCPQYRGLFL 906
Qy      679 DECFMPELVYVAIVHALLSPSSRIVLVDVHVGIFIDFGQTSANMP-----LVRDVHQ 731
Db      907 DGLMLHPCGVNFLVMSLCSBAFYVDTOQIFPIN--RVATFPYPKHLSQLGVDAVB- 962
Qy      732 CRRRTNQRKCPADVATTFPQSLYRG--CTTSGCVASISH-----VAPDYNSQA 782
Db      963 -TRRT--TLRCPADI--TFILNQYEGVMTCTSSVTRSVSHVILQGAAMVNFVSPKLG 1016
Qy      783 QTLCTFOEBS--RHGAGAMTVHEAQRTFASVLIHYNSTAGBOQLAEKSHLLVGT 839
Db      1017 KYITFQSDKSLLSNGYEDVHTVHEQSTFEDVSLVLTTPVQIISKQSHLLVSL 1076
Qy      840 RHYN-----HLYIRDP- 850
Db      1077 RHTRSIKYTVLDAVAVSLRDECVSYLLDMYKVDSVTOQLQIESVYKGVALLFYAAEK 1136
Qy      851 TDDI-----EROLN--HSAKAEFTDIPAPLET--TVY 879
Db      1137 TDDVSDMOYYDKCLPGNSTIINEYDAVTMOQIRENSLVNVDVCLMDSKVPLPRESETTL 1196
Qy      880 KSEEVQRNEVMATIPQSATP--HGAHLRLKNGFDQPDCCGVALLATGYEVFGA--R 934
Db      1197 KP-----VIRTAEKRRKPLLENLVAMIRGNNSPELVGVNIDETASLVVDKFPD 1248
Qy      935 AKINVELAPDATPKPHRAFQEGVQVVKYTNASNKIQALQTLISRYTKSADLPLHBAKE 994
Db      1249 AVLIKEKKPKKPIPLSRASLE--RWIEKQEKSTIQOL-----ADPDFID-- 1291
Qy      995 DVKRMNLSDRHMDVTEDBARAVFETQKFTQCGVYEDLLEFDDPIRIRIDFLMKT 1054
Db      1292 -----LPADVQY-----RHMIKQOPK--QR-----LDLSIQT 1316
Qy      1055 OOKVSPKPINTGKVGQGLIAHSSKLNFLVLA--AMIRILBELIRT--GSRTVRYNGPLDER 1111
Db      1317 EYVAL-----QTIYHSSKINALPGVPVSELTROLETITDSRFMPYRKTPTQ 1366
Qy      1112 EAMLEAKINQVPHATFVSADWTEFTAHNNTSELLFAALLERIGPPAAVNLFRERCGK 1171
Db      1367 EEPFSDLDGN--VP--MDILELDISKYDKSQNEPHCAVEYEWIKRLGLDDFLAEBWKGHRK 1424
Qy      1172 RTLRAGLGSVEVDGLL-----DSGAAMPCKRNTIFSAV--MLTLFPGVKFAAFGGDS 1224
Db      1425 TILKDYTAG--IKTCLMYORKSGDVTTFIGNTIIIAACLSMLPWERLIK--GAFGDDS 1480
Qy      1225 LL 1226
Db      1481 IL 1482

```

RESULT 6
 J01896
 hypothetical 191.1K protein - raspberry bushy dwarf virus
 C:\Species: raspberry bushy dwarf virus
 C:\Date: 30-Sep-1993 #sequence_revision 30-Sep-1993 #text_change 17-Mar-2000

```

C:\Accession: J01896
Rizigler, A.; Natsunaki, T.; Mayo, M.A.; Jolly, C.A.; Murrant, A.F.
J. Gen. Virol. 73, 3213-3218, 1992
A:\Title: The nucleotide sequence of RNA-1 of raspberry bushy dwarf virus.
A:\Reference number: J01896; MUID:93107856; PMID:1463559
A:\Accession: J01896
A:\Molecule type: mRNA
A:\Residues: 1-1694 <ZIB>
A:\Cross-reference: GB:851557; NID:9262149; PID:AB24599.1; PID:9262150
A:\Genetics:
A:\Map position: segment RNA-1
A:\Superfamily: cucumber mosaic virus RNA 1 protein
C:\Keywords: RNA replication

Query Match      3.3%; Score 294; DB 2; Length 1694;
Best Local Similarity 18.4%; Pred. No. 6.4e-09;
Matches 306; Conservative 228; Mismatches 589; Indels 540; Gaps 76;

Qy      47 YLRKGGTLPPTQPILAGHORV--AEVLAHPARGSTVLEIGSLHSLALHGAHPAP 105
Db      191 YELKFTQVNDGP--HNMMAAHLRLTHTDLISNFP--ADAPILDIGNWFSEHF--YGRSNV- 245
Qy      106 VADYHGCTKYGRDGRH--ITALE-----SRVATGRPEF- 139
Db      246 ---HCCPMLDLRDERHTHLMTESLSLSLRHYAGTIDLDPDALHSRKVSDSKKEFY 302
Qy      140 -----KADASLLANGIA--SRPCVDVG--SCAFKSRV 169
Db      303 KMAVHPDLILVLGIDHDSLSLCHHKFGVSNMDVLMSERRNCLTVPEECFPKAKY 362
Qy      170 GIANSLSLDVTLSELANAPKNGHLMVRAFMPEBELLYMDNVNVAELCYRHVILBEPA 229
Db      363 AITMVSIGVDLPKELIGGVGHVGLGTMADPAMLV-----ATSGY----- 406
Qy      230 VKDCAFQGGDLALHPBELDFINESQERRIERLAAGSVSRVAFPSGDDDMGDAYLHPH 289
Db      407 -----IPALRCMWSKGIWFSRFDSTMGTRHDWE 438
Qy      290 TWLAAYLVNRYPTPGFSIAHEVOR--RHGSIELRITRAPPDRLMLAVVPRTSGGLC 345
Db      439 VYSKTLT-----STVWSCGHHFVMERDKYRHGV-----LEYSIKCSGLR 480
Qy      346 RLPNIFY-----ADASGTE-----HKTILTSQKXNML----- 374
Db      481 KGDHFFHNAHEMYDKYIMKPLVVKVNDLIGDGSVECSWRREVMSKLDVRAIVECL 540
Qy      375 -----LNFQTRPKBELVDMTVLMSPARARLAIYVASEVSSNNISPADLVRTVSLY 429
Db      541 RGVKRINFGCDDAAMNDRLRIIOHMLSHSGTLVNGSTIRBEAIPPKDSPVSVTIY 600
Qy      430 VLIHIERBRAAVAK--TAKDVPGETSPWS-----LKHVLGSCC 468
Db      601 FELLITRYKESLSLAFHAGLDPDFLGS--WVSLKRVPRILGFRANLLKYVNLALF 657
Qy      469 GLRLKGTQVVTYKRVYKRVHSLGDIICVRLSPQVQV--FLPSRVPARVPHREBLE 527
Db      658 RCRD--KVSQMEFVKAVERKLTV-----LENTYIGKGLMDCPTLKEVDSAFEN 705
Qy      528 VLREAG--CYNERPV--PSTPP----- 545
Db      706 ILENVGNELPNNSSIDSGRPETPEVTMGNNPAAVLAIAISYCAVADRIGKCCERTILAY 765
Qy      546 -----VEEPQSPDAL--WHATA-----ASLPEY----- 567
Db      766 QATNGCGYIANDTVNGVFPDKMTSWQKPEBDFHFGMDGSSFIKLSMFGKIPDFGVYL 825
Qy      568 -----RATLOAGINTVQKLTLENAKTIDGLTSLSVRGLMBRGPGSGGCTLLA 621
Db      826 VTTDGTGVTTNKLKFSQVATIPATVPTTILKVDGV-----GGKXTIEYR 871
Qy      622 ALAAGKALVYAPTELEAMDRRIKPPSASATQVALAILRATAGAPATVAVIDBC 681
Db      872 RTRPG--ILILSVCANVDIIRKLAAVDSKPIRTVDSYLL-----SPSVTSGCDL 921

```

```

Qy 682 FM-----PFLVVAIVHALSPSSRYLVGDVHIGIFIDQGSANMPL-----VRDVVK 731
Db 922 FIDYEGLSHGILLALH-ISGRKVTLLFDSEIQPCN---RLADPPLKNSVEDGLN 977
Qy 732 CRBRTFNQTRCPADVATTFPOSLY--PGCTTSGCVASIS-----HVAADYRN 780
Db 978 FDEIRSTTRCQODI--TISLQMYTKTKPKIVSTYESSITIKPISSEPIPLPNAFDG 1035
Qy 781 QAOQLCTQOEKS---RHAGEG---AMTVHEAQGTFAVILHNGSTAEOQLAEKS 832
Db 1036 PVLVIQMTKDESLKLRMAKENISSEVETVHAAGSLYKNVY-YFRLTRTDNDLYTKRK 1094
Qy 833 --HLVGTITRHTNHLTRNDPTGDEROLNLSAKAEVTDI-PAPLEITTVYRSEV----- 885
Db 1095 LPHVLVAISHRTDI-----VYCTTKPEDSDPSLSAKNTITISRLTGEAS 1142
Qy 886 -----QNEVWATIPQGSATPHGAIHLLRKPFQBPDCGVALAKTGEVFG-- 932
Db 1143 GSESSYAVVFNESSEVATRPVCEVCAEM---NFPVSSD---ALYQKEVPIYGA 1195
Qy 933 ---GRAKIN-----VELAEPDAPKPRAPQE-----GVQW--VKVTN-- 965
Db 1196 PDPKGRASVYNGSVIRAIIEELTPGNTSIDTDLDELVEVGPMSLVGSIKIMDVSKISPL 1255
Qy 966 -ASNK---HQALQTLISRT-----KSSADPLHAKEDVYKMLN-SLDRHWDM 1009
Db 1256 FTNNKRAVPHLPFGALLRRNTSSRQVGAIEKRNANVNSQKFDLENLANKAVERPFD 1315
Qy 1010 TVTEDAADR-----AVPETOLKPTORGATVED-----LLEPDDPY-----IRDIDPLMK 1053
Db 1316 FIDMEKRSKLPVGLSSAEOIQTYONKTKNTDPCVALISPIQYKEMIKRDKVKNLT 1375
Qy 1054 -----TOQKVSPEKPINTGKVGQ3---IAHSSKINFLVLAAMRILEILRT 1096
Db 1376 DGAQSEYTKAATITVHPEITQVATAIFGQFKTRLACRKPLNIPL----- 1422
Qy 1097 GSRTVRSNGLPDEEEMLEAKINOVPHATFVSADMTEDPTAHNTSELFPALLERIG 1156
Db 1423 ----BHNDLISGLTKYHLGSENN-----TFTEIDSKPKDQSGEIHQIIDLKFG 1472
Qy 1157 TPAAVVLPFRBCGKRTLRAGLG-SVEDGLDLSGAAMTPCRRITISAAVMLTF----- 1211
Db 1473 CDEPFAVLMSTARRSSIFPDQNGIGFKTPQRRRTGAFPLGSLTAMLAFAVISDDP 1532
Qy 1212 -RGVKAFAFGKDDSLLCGSHYLRFDASRLHMGERYKTKHLKVEQKIVPIYG---LLVSA 1267
Db 1533 RKIRIRVWLVGDDSLICS--YGRIVQVLEPLGLDIFMNS-CKL-VQPAQPFYASRYILRG 1588
Qy 1268 EGV--VLDPVSALKIFGRCYTSLLYSKYVEAVRDTIKGMSDARHSLCHMSACYNY 1325
Db 1589 DELICVDPDPYKYLKYL-GR-----KQVPDN-----QASICEIR---TG 1622
Qy 1326 APESAAYIIDAVVRFQGGDFPFGQLRV---VRAHVQAPDAYSS 1365
Db 1623 LADSAKITFDIVK-----QKLALIVQRYNKAAPSLYDA 1657

```

RESULT 7

```

J02144
183K protein - tomato mosaic virus (strain Ob)
N/Contains: 126K protein
C/Species: tomato mosaic virus
C/Date: 03-May-1994 #sequence revision 03-May-1994 #text_change 16-Jun-2000
C/Accession: J02144; J02143; J02158; J02157
R/Padgett, H.S.; Beachy, R.N.
Plant Cell 5, 577-586, 1993
A/Title: Analysis of a tobacco mosaic virus strain capable of overcoming N gene-mediated
A/Accession: J02144
A/Molecule type: mRNA
A/Residues: 1-1616 <PAD>
A/Cross-references: GB:L11665

```

```

A/Note: this form is translated based on a read-through of the codon TAG for residue 11
A/Accession: J02143
A/Molecule type: mRNA
A/Residues: 1-1115 <PA2>
A/Cross-references: GB:L11665
A/Note: this form is translated based on the interpretation of the TAG at position 116
R/Ireda, R.; Watanabe, B.; Watanabe, Y.; Okada, Y.
J. Gen. Virol. 74, 1939-1944, 1993
A/Title: Nucleotide sequence of tobamovirus Ob which can spread systemically in N gene
A/Reference number: J02157; MUID:93389450; PMID:8376970
A/Accession: J02158
A/Molecule type: genomic RNA
A/Residues: 1-152, 'K', 154-872, 'N', 874-1616 <IKR>
A/Cross-references: GB:D13438; NID:9436229; PIDN:BA02700.1; PID:9436230
A/Note: the codon TAG for residue 1116 is translated to Trp
A/Accession: J02157
A/Molecule type: genomic RNA
A/Residues: 1-152, 'K', 154-872, 'N', 874-1115 <IKR>
A/Cross-references: DDBJ:D13438; NID:9436229; PIDN:BA02700.1; PID:9436230
C/Comment: This protein is involved in replication of the RNA genome.
C/Superfamily: cucumber mosaic virus RNA 1 protein
F/1-1616/Product: 126K protein #status predicted <MAT2>
F/1-1115/Product: 126K protein #status predicted <MAT2>

Query Match 3.3%; Score 291; DB 1; Length 1616;
Best Local Similarity 19.4%; Pred. No. 8.9e-09;
Matches 327; Conservative 220; Mismatches 671; Indels 464; Gaps 70;

Qy 24 QORAVKLDPAFPAKALETL--HRLY--YPLRFGGTLPTQHPILAGHQRAEVLNFA 79
Db 42 KDRRPKNPFSKINEGQTLVSGAYPEFQTFNTQL--AVHSLAGRLSELEFLYLMQV 99
Qy 80 RGRSTVLEIGPSLSALAKLHGAAPVADY-HGC-TKYGRDSSRHITALSBSVATGR- 136
Db 100 PYGSLTFYDIGNPAHL-FKG-----RDVHCMPRLDRLDRHENGQDSVATYLSRL 152
Qy 137 -----PEFKADSLANIASRTFCVGVSCAKSR-----VGIANHSLYDYLE 182
Db 153 NADNKVLPAPQBARFQYRSRSDVVC-NNTFOCCSNRSGGRVYALSLHSLYDIPAD 211
Qy 183 ELANAEHNGLAHVRAFMHPEBELLYMDVNVNLELGRFHVIEBPMAVNCACAGGDLRL 242
Db 212 ELGALLARKVHTLYAFAHELLLEVTY-----ELPTIGITPSRDD-KI 258
Qy 243 HPELDFINESQERRIERLAAGSYSR---AVISGDDWDGAYLDF-----HTWL 292
Db 259 NF---CFSNES-----TANSHSYSNLKYCKTFYPASNRE--VYKKEFFLTRVNTWF 307
Qy 293 -----ATLYLR-----NYTPFPGLHIEVQRHSGSIELRT 325
Db 308 CKFTKLDYTYTLRYGVHRGDDQGFYSAMEDAWHYKTKTAMNSERIVLEHDSVNYWF- 366
Qy 326 RAPDGRMLAVVPRTSQGLCRIPNIFYADASGTEHTITLTSQHKVMTLMPQTRPEKE 365
Db 367 ---PKKDMVIVLPVSL-----ETQRTKKEVIVSDPPTVTLNHRITQAKA 413
Qy 386 LVDMTVLSPARARLAIIVASVTESSWNISPADIVRTVVSILYHLIIR-----RR 438
Db 414 LTYNNVL-SFVBSIRSRVITINGTARSMDVDYKALLQSNAMTFPLTKLSMLDELVSX 472
Qy 439 AAVAVTAKDDVGE-----TSFREL--KAVLGSCGGLR-NLKTQDVVFTKRVV 485
Db 473 FTLSAKSVHEHVDKIRGGGNMFPILKSLARKKILSGSABEELIEVPMYTFHDFRV 532
Qy 486 DKTRYV-----HSLGDI-----ICDVR----- 501
Db 533 AETKASVEMPTIIDSGLSABSYALSLSLSTLNSKQDLKPKRMCAINCNPDI 592
Qy 502 -----LSPEQVFLPSRVPP---ARVFRDELEVLRAAGCYNRPPVSTPPV----- 546
Db 593 KIIVAVLVSNSGVTLPFKBPTBGNMABAKSGKDEVLTLGSGTDNTDLTSKSMVSGSL 652
Qy 547 -----EEQGFADLMHATTA-----SLPEYRA 569

```

```

Db      653 PLGIAISEISCTPPRNEEINSLER--YMLAAESVSNKMASIVSGPLQVQOMONTYD 710
Qy      570 TLQAGLNTDVYKQJ-----KITLEN-----ALKTIDG 595
Db      711 SLAASIASAVSNLKKLVKDSVGFQDSLSKVGIVDKKMMMLKPLTKHNSGVVQKFDG 770
Qy      596 -----LTSVPRGLEM 606
Db      771 KCFLLALSYHNEPLICDADMSKVAVSNESMVYSDMAKLEVLKRSIGEMPISSVSAK-VYL 829
Qy      607 YGPGPSGKTGTLIALAAGKALVAPTRBELREMBRIRKPSASATQH-----VALA 661
Db      830 VGVGECCKTEILKRVNFS--EDLVLVGKEAAMIRKKAOSGIYANNNDVKTVDSP 887
Qy      662 ILRRATGAGAPPAVVIDECFMPFLVYVAIVHALSPSSRIVLVGDVHQIGTFIDPGTSAN 721
Db      888 LKNLKGKPGVCFKRLFVDEGMLHPGCYFVLVTLSLCNEAFVFGDTQQLPYIN--RVGN 944
Qy      722 MELVADVYQC-----RRRTNQTACGPAVDVATTFGSLYFGCTTSG-----CY 767
Db      945 FFPQHPFSLIVDETEKRT--TLRCPDV--THFLNOCYGAATVTTKTORSVGLERV 999
Qy      768 ASISHVAPDRNSOAGTLCFTQEBK-----SRHAGBAMTVYHBAQGRTPASVILHYNGSTA 823
Db      1000 GGAANVPVTKPLKGIYVFTOSDKLTMLSR--GYQVNTVHEIQGETYEVSIVRLTTP 1058
Qy      824 EOKLAESKSHLVGITRHTN--HLVIRDP--TGDIERQLNLSAKAEVFTDPA-PL 874
Db      1059 IHIISSPHVVLGILRHRCFRKYVTVVLDPLVKLVRLDECVSNFLDYVMDVSVAWQL 1118
Qy      875 EITTYKPSSEVQNEVMATIPPOSATPHGAHILRNKFGDOPDC-----GCVAAKTGYE 929
Db      1119 QVSGVYLAEN-----LFVQAPKS-----GDAODLPQYDCKLPGNSTVLA 1158
Qy      930 VVG-----GRAKINVE--LABPDATPKHAFQEGVQVQVYKVTASN--KHQALQTLISR 979
Db      1155 EFDATVANCSDISLANKDCVLDPSKSVPLPRDNTKQPMTPVIRTAHERRSQGLENLVA 1218
Qy      980 YTKRSADLPLEHAKEDVKEMLNSL--DRHMDWTVTEDAR-----DRAV----- 1020
Db      1219 MKRNPNSELPESGTVDMENTASVADRPFDSYFLKDKLSGCSLGDGSGKNIIDROLIRM 1278
Qy      1021 FETQTKTORGCTVEDLEBPDPYIRIDIDFLMKTQ--QVSPKPINTGVGCGIAHSGS 1078
Db      1279 MEKOEKSTI--GQLADYDFVDLPAIDQYHIIKSQPKQKLDLSIOSEYPSLQTIYVHSKK 1336
Qy      1079 LNFVLAAMRILEILR-----TGSRTVRSNGLPDEEAMLEAKINQVPHATVSD 1132
Db      1337 IN--ALFGPIFSELROMLSAIDTSRYLFTFKTPEQIEFPSSDADAQPMEEV--LEID 1391
Qy      1133 WTEPFTAHHNTSELLEFALLERIGTPAAAVNLFRERCGRRTLRAKGLSVEVDGL----- 1188
Db      1392 VSKYDQSQNEHFCAVYEIWKRLGIDELFABVWKQHRKTKTLKDYTAG--IKTCMWR 1448
Qy      1188 DSGAAMPFRKNTIFSAANWLTFRGYKF--AAFKGDSLL-----CSGHYLR 1233
Db      1449 KSGDVTFPIGNTVVIIAACMASMLPMEKVIKAAFCGDSLSVYLPGKCELPNIOSCANLMMN 1508
Qy      1234 FASRL-----HMGERTKTKHKEVQKIVPYIGLVSABOVVLDVFRSALKIFGSCYNS 1288
Db      1509 FAKLFFKTYGYFCGRYVVIHDBGAI-----YVVDLKIISKLGAKHITD 1553
Qy      1289 ELLYSKYVAVDITKGSMDARYSHLLCMSACYYNVAPEASAAYITDAVVRGGRGDPPE 1348
Db      1554 KEHLERFRI SLADVSLSNNCAVYAOQLDAVREHVHTAP--PGSFVYKCIYKFLSNRVLPE 1612
Qy      1349 QL 1350
Db      1613 SL 1614

```

RESULT 8

```

NM18
183.3K protein - tobacco mosaic virus
C:Species: tobacco mosaic virus, TMV
C:Date: 17-Dec-1982 #sequence_revision 17-Dec-1982 #text_change 15-Oct-1994
A:Accession: A04194
R:Goefet, P.; Lomonssoff, G.P.; Butler, P.J.G.; Akem, M.B.; Galt, M.J.; Karn, J.
Proc. Natl. Acad. Sci. U.S.A. 79, 5818-5822, 1982
A:Title: Nucleotide sequence of tobacco mosaic virus RNA.
A:Reference number: A93926; MUID:83299880; PMID:6964389
A:Accession: A04194
A:Molecule type: genomic RNA
A:Residues: 1-1615 <GDB>
A>Note: The authors translated the codon AUU for residue 1302 as Tyr
A>Note: this sequence contains proteins 125.9K (residues 1-1116), 87.4K (336-1116), 83
A>Note: readthrough of the terminator codon UAG between codons for Gln-1116 and Gln-11
A>Note: variant forms are found with 9-Ser, 14-Glu, 21-Thr, 37-Asp, and 683-Ala
C:Superfamily: cucumber mosaic virus RNA 1 protein

Query Match          3.1%; Score 278; DB 1; Length 1615;
Match Local Similarity 20.0%; Pred. No. 5.3e-08;
Matches 312; Conservative 193; Mismatches 580; Indels 474; Gaps 67;

Qy      24 GORAVKLPAPPLKALFTL--HRLYPLRFKGTLPPTQHP--LAGHQRVAB----- 72
Db      42 RRRRKVPKSKYISSEQTILATRAYPEQI--TYNTQNAVHSLAGLSLELYLMMQ 98
Qy      73 -----EVLNFA-----RGRSTYLEIGPSLSHSLKLGAPNAPVADYHGCTKXYGRDQ 120
Db      99 IPYGSITVDIGNFGNFASHLFGKRAYV-----HCCM-----PMLDV-----RDI 135
Qy      121 SRH-----ITLERSVAIQR--PEFKAD-----ASLLANIARTCTVDGVCAPFS 167
Db      136 MEHQGQKSTELVYSRLRGGTVPNFQKAPDRYAKLPEDAVCNHTQTQTMHQPQOSG 195
Qy      168 RV-GIANSLVDVLEELANAFENHGLHNVRAFMNPEELLYMDVNVVAELGRTFHVLE 226
Db      196 RYVALAALHSIYIDIPADEGALLRKNVHTCYAAFPSENLLDESYVLD----- 245
Qy      227 PMAVDCAPQGGDLRLHPELIDFINESQERIRIELRANGSYR-----RAVIFGDDDWG 281
Db      246 --EINACFSRDD-KLTF--SFASBS-----TLNCHSYENILKYCKTFFPSAN--R 291
Qy      282 DAYLHDF-----HTML-----AYLYR-----NYTPFGFSLH 309
Db      292 EYVMEKFLVTRVNTGPFCKPSRIDTFLLYGVAHKSVDSEQFYTANEDAMHYKTLAMCNS 351
Qy      310 IEVQRHSSSIELRTTRAPGDBMLAVVPTSGCLRIPIRYVADASTGTEKTLITSOH 369
Db      352 ERIILDESSVAVYWF-----PKRDMVIVPLDISL-----ETSKRTKEVYLVSKD 397
Qy      370 KYNMLANFMQTRPEKELVDMTVLMSFAPARLRAIVVASBVTESSWNISPADLVRTVSLY 429
Db      398 FVFVTLNHRITVQAALATYANVL--SFVESISRIRVIINGVTASBWDVKS--LLQSLSMTF 455
Qy      430 VLIH-----IEBRBAVAVKTKADYFGSTF-----WESLK-----HYLQSG 467
Db      456 YLHTGLAVLKDDLISKFSLSKTYCQHVMBEISLAFGNAPFSVBRLLNRGLIRVAGDA 515
Qy      468 GGLRNMLKGTVDVFTGRVVDKYRVHSLGDIICVRLSPQVGLPFRVPPARVPHREBLE 527
Db      516 LEIR--VPDLVYTFHRLVTRYKA--SVDMPALDIRKQMBETVM-----YNLSBELS 564
Qy      528 VLRAGCTN-----BRP----- 539
Db      565 VLRSEKDPDVDFVFSQMSQSLVDPMTAAKVIIVAVNSNSGLTLTERPTEANVALAQDQ 624
Qy      540 -----VPSPPVPEEQ-----GPAD-----LWH-ATA 561
Db      625 EKASGALVTVSRSEVBSMGSMARGELQLAGADHPRESSYKNEBISLEIQFHMAVA 684
Qy      562 ASLPEYR-----ATLQAGLNTDVYKQLKITLENALKT----- 592
Db      685 DSLIRKQSSIVYTPIKVQCKMKNFIDSLVAVSLSAVSNVYKILDTAIDLETQKQGV 744

```


QY 593 -----I 593
 DB 745 LDVASKMLIKPRAKSHAMGVETHAKRHVALLIEYDQGVTCDDMRRAVSSSVYS 804
 QY 594 DGLTSPVRL-----EMEGPPSGKTGTLAALBAAGKALVYPTBELR 640
 DB 805 DMAKRLRLRLRNGEPHVSAAKVLVDVPGCGCKTEILSRVNF--DEDLILVPGQAA 862
 QY 641 EAMDR-----IKPPASATQHVALLIRATAAGAPATVVIDECMFLLVVAIVHAL 695
 DB 863 EMIRRRANSGLIATADNVKTVDSPFMNFGKSTRQFKRLFIDEGMLHTGCVFLVAM 922
 QY 696 SPSPRIYLVGDVHQIGRID----PGTSANMLYRDVVKOCRRPTQTCRCPADVAT 751
 DB 923 SLCEIAVYDQDQIPYINRVSGFPYAPFAKLEVDEVE--TRRT--TLRCPADV--TH 975
 QY 752 FPGSLVPG-CTTSGCAVASH-----VAPDRNSQAOQLCTOBEKS--RHGAEG 799
 DB 976 YLMRRYGFWMSTSSVSKSVSGENVGAAYINPISKPLHGKILFTQSDKALLSRGYSD 1035
 QY 800 AMTVHEAQRTFASVILHNGSTAEQKLAEKSHLVGTRHTNL---YIRDPGDI 855
 DB 1036 VHTVHEVQGETYSDVSLVRLPTPVSIAGDSPHVALSRHTCSLKYTVVMDPLVSI 1095
 QY 856 RQLNHSA-----KAVFDPIDPAPLEITTYKPESEVQRNEVATIP----- 895
 DB 1096 RDLKLSYLLDMYKVGATQ--QLQIDSV-----FKGSLFVAAPRTGDISDMQFYD 1147
 QY 896 ---POSATPHGAHILRKNFQDQPCGCVALAKTGEVFGRAKINTELAEPTAPKPR 952
 DB 1148 KCLPGNST-----MANNF---DAYTMLTDSLAVKXCITLMSVAAKPKQIKPL- 1195
 QY 953 AFQGVQVQKVTASNGHQAQLTLLSRYKRSADLPHEAKEDYKMLN--SLDRHMDWT 1011
 DB 1196 ----IMVRTAAMPRTQGLLENLVAMIKRNFAPFELSGIIDIENIASLVVDKFP 1250
 QY 1012 TEDAR---DRAVF-----ETQLKFTORGSTVEDLEDPPIYRIDFLMKTQ--Q 1056
 DB 1251 LKEKREKKNKVVSLRESRLNRMLEKQOVIT--GQLADPDPFVDLPAYDQYHMYKAPKQ 1308
 QY 1057 KVSFKPIINTGVGGGIAHASKSLNPLVLAAMIRILEILR-----TGSRTVRYNG 1110
 DB 1309 KLDTSITETPALCTIYHSHKINATGPR--LFSRLRQLDSVDSRPFEPFRKTPAQ 1355
 QY 1111 EEMALTEAKINQVPHATFVSADMTPTAHNNTSELPAALLERIGTPAAAVNLFRRCG 1170
 DB 1366 IEDPFGDL-D-SHP-MOVLIEDISKYXQNEPHCAVEYELWRILGFEDPLGEVWKQGR 1423
 QY 1171 KRTLRAGLG-SVEVDGLDSGAAMPCCRNTPISAAVMLTFRGVK--AAFKGDDSL 1226
 DB 1424 KTLTKDYTAGIKTCIWMYQKSGDVTTFIGNTVIIAACLASMLPMEKIKGAFCDSDSL 1482
 RESULT 9
 WMTGM 183K protein - tobacco mild green mosaic virus (strain U2-TMV)
 N/Contains: 126K protein
 C/Species: tobacco mild green mosaic virus
 A/Title: host Nicotiana tabacum cv. Samsun (tobacco)
 C/Date: 30-Jun-1992 #sequence_revision 30-Jun-1992 #text_change 23-Jul-1999
 C/Accession: A35520
 A:Reference number: A35520; MUID:90320127; PMID:2371769
 A:Accession: A35520
 A:Molecule type: genomic RNA
 A:Residues: 1-1608 <SOL>
 A/Cross-references: EMBL:M34077; NID:G335243; PIDN:AAA47934.1; PID:G335244
 C/Comment: This protein has RNA polymerase activity.
 C/Superfamily: cucumber mosaic virus RNA 1 protein
 F:1-1111/Product: 126K protein #status predicted <RP>

Query Match 3.04; Score 268.5; DB 1; Length 1608;
 Best Local Similarity 19.24; Pred. No. 2e-07;
 Matches 318; Conservative 212; Mismatches 563; Indels 565; Gaps 78;
 QY 24 QQRAVKLDPAFPKALETL--HRLYYPLRFKGGTLPPTQHPI--LAGHGVAR----- 72
 DB 42 RDRRPKNFKTISEBETTLVSNAYPEFQ---TFRTQVAVHSLAGLALLETLMQ 98
 QY 73 -----EVLNPA-----RGRSTVLEIGSLHSALKMGAPNAPVADYHGTCKYGRDG 120
 DB 99 VPYGSPTTYDIGNFAALFPGROYV-----HCCM-----PILDI-----RDI 135
 QY 121 SRHITALESSVATGR-----PEPKADSLANGIASRTFCVGVGCAFR----- 166
 DB 136 MRHGGQDSLEMTLSRLSRSNKVIPEFQRAAFNRYPAPPEVCCSKTPQDCRIHPENSG 195
 QY 167 SRGVIANHSLYDTLLELANAFENHGLHWRAFEMPEBELLY-MDNVVAELGFRFAYIE 225
 DB 196 RRYAVALHSLYDIPHEFGAALLSKNIVCYAASILAALLDQTEVTLNIEIGATF--- 251
 QY 226 EPMAYKCAFQGGDLRLHPELDFINESQ--RRIERLAAGSY--SRRAVIFSGDDDMGD 282
 DB 252 -----KRGDVSFPFADBSLTLYSHKYKNIILHYVVKSYPPASGRIVYFK-----E 297
 QY 283 AYLDHFTTL-----AYLVN-----YTPPGFSLIH--EYQ--R 315
 DB 298 FLVTRVATWPKCKFTKVTYTLLYKSVQGVGDSQDFEAMDEAPYKKTLMFNTERRAIFR 357
 QY 316 HGSSIELRLIRAPPGDEMLAVPR-----TSQGLCRIPNIFYAASGTEHKTILTSQHK 370
 DB 358 DTASVNWPF-----PKAKMVIYPLFGSITSSKMT-----SEVTVRDF 398
 QY 371 VNNLNMQTRPEKELVDMTVLMSFAPARLYAAVVAEVTSSKNISPA----- 419
 DB 399 VYTVLNIIRYQAKALYQVVL-SVESHSRYVINGVTARSEMDVDKALQPLSMTPFL 457
 QY 420 -----DLV-----RTVSLVYLHITERRA 439
 DB 458 QTKLAALODDIWAGKFCCLDKTTSBLMDVGRFGNVPPTIKERLVSRLILDSEN--- 514
 QY 440 AVAVK-----TAKDVFGSTSPWESLKH-----VLGSCGLKNTKGT--- 477
 DB 515 ALKIKIPDLVYTWKDFVATYTSBELPHLDIKDLSEABQMDALSELITLKADNPDI 574
 QY 478 -----VVF-----TRVVDKYVHSLGDIICVRL 502
 DB 575 AKPKDMCKALDVSDDVAARYIVAAENRSGTLTFDKTEENAKALKSTASERVCLER 634
 QY 503 SPBOVGF-----LPSRVPPARVFN-----DREBLEVLRBAGCTNERPVST--- 543
 DB 635 TSBEVNVNKEFSIAEKRLPYCASHGLTNANLHIOBLESLND---FHKACVDSYITQMA 691
 QY 544 -----PVVEPQGFDD-----LMHAT----- 560
 DB 692 SVYTTSLKLYQMKNVYDSLAAISLATVSNLCSLKQDVGYSDSREKVGMDVTLKMTL 751
 QY 561 ---AAS-----LPEYRATLOAGL-----NTDVOKLKIT-----LENAL 590
 DB 752 LKPAKSHSGVULVDYGMKFTALLSYBGBRMVTSBWRVAVASDPMYVSDIAKLQNL 811
 QY 591 KTI-DGLTSPVNGLEYSRPPSGGKTGTLAALBAAGKALVYAPTRBELRANDRIRK 649
 DB 812 KTMKDGPHSPFTAMVVLVDVPGCGK-----YKGDPERPDLDBDLILVP 855
 QY 650 PSASATQHVALLIRATAAGAPATV---VIDECFMP-----LVYVAI 691
 DB 856 GKQAAAM-----IRRRANSGLIRATMDNRYTDSGLMHPKPSHKLFIDEGLMLHTGC 910
 QY 692 VHAL---SPSPRIYLVGDVHQIGRIDPGQTSANMP-----LVRDVVKOCRRPTQTCR 741
 DB 911 VNFLVLSGDDIAYVGDQDQIPFIN--RVQNPFFYKIHFKVQYDVEN--RRT---TL 962

QY 742 RCPADVATTPFQSLYPGCTTSS-----GCVASISHVADPYRNSQAOTLC 786
 DB 963 RCRGDU--NFIQSLKEGAVTTTSTYQSRVSSMIGKVLNSV-----KELKRIYT 1014
 QY 787 FTQEEK--SRHGAEGAMTVHEAQGTTPASVILHYNGSTABOKLAEKSHLLVGTIRHTN 843
 DB 1015 FTQADKELEBEKGYKNNVTYHBIQGETFEDVSLVRLATPDLTILSKSPHVALFRHTK 1074
 QY 844 ----HIYIRDPDIDIERQLNHSKAEVFDIPA---PLEITTVKP-SEEVQNEPMAT 894
 DB 1075 SFERYVVVDLPL-----VDIISDLSLSEFLLEMYVVEAGSHQLQMDAVFKGH 1122
 QY 895 PQSATPHGAHILLRKNFGDPDPCG-----CVALAAT--GYEVFGRAKIN----- 938
 DB 1123 NLFVALPKS-----GDFPLOFYDVCLPQNSTILANKYDAVMTKLBDNSLANKDCV 1173
 QY 939 VELAEADATPKPRAFQEGVQWVKVTNASNKQALQTL--LSRYTYSADLPLHAKEDV 996
 DB 1174 LDFSKSIIPMKEXKPCLEPV---LRTAAEPRAAGLLEVLVAMIKRNFAP----- 1221
 QY 997 KRNLSLDRHMDVTYEDARDRAVFTQKFTQRCGTV-----EDLLE--- 1039
 DB 1222 -DLTGDIESTASVAVDVFDSYFTKKEKYTKNIAGVMTKOSMMKLENREKSVLLDDLA 1280
 QY 1040 ----PDPPYRIDIDFLAKTQ--QKVSPPKPIGTGKVGQGLAAHSKSLNFWLAAMIRILBEI 1093
 DB 1281 NNFTDLPAIDQYKMHKIQKQKDLSTONBEPALQTIYHNSKQINGILAGSELTRLL 1340
 QY 1094 LRT--GSRIVYSNGLPDEBEAMLLEAKINQVPHATFVSADWTEPTAHNTSELLPAAL 1151
 DB 1341 LEAFDSKKLFFETRKTPEDQIQEFPFSLD-SHVP-MDVLIELISKYDKQNEHFCAVEYEI 1398
 QY 1152 LEAIGPRAAAMVLFRRRCGRTRLRAKGLSVEVDGL-----LDGAAATPCRNITFSA 1204
 DB 1399 WKRLGLNEFLAEVWKQSHRRTTLKD-----YTAGIKTCLMYQKSGDVTTFIGNVTIJA 1452
 QY 1205 AV--MLTLPRGVKFAFKGDDSL-----CGSHYLRFPDAS--RLHMG--B 1243
 DB 1453 ACGSLMLPMKVKIK-GAFCDSDSVLFFPKGLDPDIDQSCANLMMNEBAKLYRKRYGFCG 1511
 QY 1244 RYTKKHLKVEVQKIVPYIGLVSADQVLDPRVSAKI 1281
 DB 1512 RYIIHHDKAI-----VYVDPLKILSL 1534

RESULT 10
 S48699
 178K protein - tobacco mosaic virus (strain cr-TMV)
 N:Alternate names: readthrough protein
 N:Contains: 122K protein
 C:Species: tobacco mosaic virus, TMV
 A:Variety: betain cr-TMV
 C:Date: 06-Dec-1996 #sequence_revision 06-Dec-1996 #text_change 20-Sep-1999
 C:Accession: S48699; S48659
 R:Porokhov, Y.L.; Ivanov, P.A.; Novikov, V.K.; Agranovsky, A.A.; Morozov, S.Y.; Efimov, F.S. Lett. 350, 5-8, 1994
 A>Title: Complete nucleotide sequence and genome organization of a tobamovirus infecting
 A:Reference number: S48659; MUID:9431372; PMID:7545946
 A:Accession: S48699
 A:Molecule type: genomic RNA
 A:Residues: 1-1601 <DOR>
 A:Cross-references: EMBL:229370; NID:9488713; PID:CA82559.1; PID:9619908
 A:Experimental source: tobamovirus infecting crucifereae plants (cr-TMV)
 A>Note: readthrough of the terminator UGA occurs between codons CAA for 1107-Gln and CAA
 A>Note: the internal stop codon is translated as X
 C:Superfamily: cucumber mosaic virus RNA 1 protein
 F:1-1601/Product: 122K protein #status predicted <PRO2>
 F:1-1107/Product: 122K protein #status predicted <PRO1>

Query Match 3.04; Score 265.5; DB 2; Length 1601;
 Best Local Similarity 18.74; Pred. No. 2.9e-07;
 Matches 319; Conservative 209; Mismatches 635; Indels 547; Gaps 69;

QY 2 YAKATDVAVRYAA-----DVA-----YANVLOO-----RAVKLDPAPRLKALETL--H 43
 DB 4 PGGTIDMQTLQAAAPNSLVNDLASRYVDNAVESELNARSRRPKYHSAVSTGTLAT 63
 QY 44 RLYPELRFKGGTLPTPOPIIAGQORVAEVLNPNRAGSTVLBIGPSIHSALKLGAEN 103
 DB 64 NAYPBEISFTHTQSAVHSLAGGFRSLEBYLQMVPQSLTYDIOGAP-SHILFKG--- 119
 QY 104 APVADY-HGC-TRKGTGSGRHITLSEBSVATGR-----PEPKADASILANGIAR 153
 DB 120 ---RDYVHCMBNLDVRIARHBEKEAHSYVNLKQOQREVPPEYORA---FNNYAN 173
 QY 154 ---TFCVDGVSACFKSRVG-----IANHSLDVTLEBLANFENHGLMVARFMMHRE 205
 DB 174 PHFACDPRPQGCBLTANGDTTYVALISTYDIVEBSGLAKKNVTCFAAFPHEN 233
 QY 206 -LLYNDVYNAELGYRFHYIEBPMAVKCAPOGGDLRLHPBELDINSQERRIERLAR 264
 DB 234 MLLDCDTYLDIEIGATFQ-----RAGDNLSPF--FANSTLN----- 268
 QY 265 GGSRRAVIFSGDDMGDAIYLDFTWLAIVLYVPPFGPSLITVQRRHGSIELRI 324
 DB 269 -----YTHSFENIITKYCKTFEPASQRFVYHKEPLVTRVNTWYCKF 309
 QY 325 TRAP-----PGDR 332
 DB 310 TVVDFTFLRGYVHNNDCEBPYKAMDAMHKTLANLMBRTFKDAAALNFWPPKR 369
 QY 333 MLAVPRTSGGLCRIPNIPTYDASGT-----BHKITLSQKRNMLNFMQTRPEKELVD 388
 DB 370 DMVIVP-----LF--DASITGRMSRRFMMVNDPVYTVLNIKITKYQAKALY 415
 QY 389 MIVYLSFARALRAIVASSEYSSMNISPADIVTVTVSLYLIHIERRAAVAKTAD 448
 DB 416 ANVL-SPEVSIISRVIINGVTAARSMDTKALGPLAMTFILTKLGHVODEITLK--KF 472
 QY 449 DVFGETS---FWESLKHVLSGCGRLNKGTDV-----VPTKCV 484
 DB 473 QKFDRTNELITSLCDALMGV--IPSVKETLVRGQPVYVAQALEIKIPELICYTPADRL 530
 QY 485 VDKYR---VHSLGDIICDVRLSPROVGLPSPRVPARVYHREBELVLEBAGCYN--- 536
 DB 531 VLQYKABRFQ---CDLSKPLBE-----SEKYNALSELVLENIDFDLBAF 576
 QY 537 -----RPVPSTPVEEP-----QGPAD----- 555
 DB 577 KTLCOQKSVDPMAKVVVAINKCELTLPKKPTEBIBISLKTGTSABEKDVLSTON 636
 QY 556 -----LMHATAASLPEY-----RATLOAGLNT---D 578
 DB 637 DAPFCVKNLVGSGVPAYGMCCKGGFPLDLDVIDFHLKSDAVAKRGTMASAVYTGSI 696
 QY 579 VQOLK-----ITLENALKTI----- 593
 DB 697 VQMKQYIDYLSASLSATVSNLCYLRDVHGVDPBSQSKGWDVARGRWLLKPAKSHA 756
 QY 594 -----DGLTSPVQGLB----- 605
 DB 757 KGVADANAKLIVILLNMDGKPCVDETFWRVAVSSDSLTIDMGKLTLLTTCSPNGEP 816
 QY 606 -----MYEGPPSGSKTGLTIALAAGKALYVPTBELBAMDRIK-----PP 650
 DB 817 ERNAKYLIVDGYFGCKTSEIIEKNFS--EDLILVPGSAKMIIRANAHGVIRADQ 874
 QY 651 SASAQHVALILRRATAGAPPAIVVIDECFMPLVYVAIVHALSPSSRIYVGDVQI 710
 DB 875 NVSTVDSFLMHPSRRV-----FKSLFIDEGMLHTGCVNFIILLSCQCVAAVYGDQOI 928
 QY 711 GPIDPGTSANP-----LVADVQCORRETRVQTRGCPADVAATTFQSLYFG---C 760
 DB 929 PFL--CRVAVPPYPAHFAKLVAD-EKEVRRATL--RCPADV--TYLANKYTDQAVNC 978
 QY 761 TTT-----SGCVASISHVADPYRNSQAOTLCFTQEEK--SRHGAEGAMTVHEAQGT 811

```

Db      979 TSAVERSVKAEVNRKGLNPTLPLEBKILFTQADRFELLEKKYQVNTVTHFVQSTY 1038
Qy      812 ASVLHNGSTAEOCKLAESKHLVIGTRHTN---HLYIRDPTGDIRQLNHSK---- 863
Db      1039 EKTIVRLTSTPLEIISRAAPHVALTRHTTRCKKYTVLDPMTNVNISEMKLSNFLD 1098
Qy      864 -AEVPTDIPAPLEITTV-----KPSSEVQRNEVMAITIPQSAITPHAIHLAKN 911
Db      1099 MYREAGIQXQLOIDAVKGNLFLVQTPSGMDMOMQRYNDTLFGNSTILNEYDAVTN 1158
Qy      912 FGD-----QPDGCVLAKTGYEVFGRAKINVELAPDAPKPHAFQEGVQVWVYVAS 967
Db      1159 LRDISLANKDC-----RIDFSKSVQLPKEGQPIFLK----PIKTRRA 1195
Qy      968 NKHQALQTL--LSRYTKESADLPHEAKEDVKRMNL-SLDHRMD-----WTVT 1012
Db      1196 EMPRTAGLLENLVAMIKNNMNAFDLTGTDIEDTSLAVEKEKWDSDYIDKEFGTNEMTT 1255
Qy      1013 EDANDRAVFERQKFTQKGTEDLE---PDDPIRIDIDPLMKTQ--QKVSFKPIVTK 1067
Db      1256 RESFGRMLSK-----QSSSTVGQLADFNVDLPVDEYKHMITSQPKOKDLSTIDEXY 1309
Qy      1068 VGGGIAHSKSLNFVLAAMIRILEBILR---TGSRTVRYNGLPDEEAMLLEAKINQVP 1124
Db      1310 ALQITVHSKKINAFGMFSELTMLERIDSKFLTYTRKTPAQIDFSDLDSTQA- 1368
Qy      1125 HATFVSADWTEEDTAHNTSELFPALLERIGTPAAVNLFRERCGKTLRAKIGASVEV 1184
Db      1369 -MEILELIDISKYDKQNEFHCAVEYKIEWEKLIDEMLAEVWKQGRKTTLDQYTAG---I 1424
Qy      1185 DGL-----DSGAAMTPCMTITPSAIV---MLTFEGVCPFAAFKGDLSL----- 1226
Db      1425 KTCMLYORKSDVDTFIGNITIIAACLSMTIPMDVKIK-AAFCGDSLSIYIPKGLDLPDI 1483
Qy      1227 -CGSHYL-RFDSRL-----HMGERYKTKH-----LKYEVOKIVYIGLVSABOVLDP 1274
Db      1484 QAGANLMMNFEAKLFRKKYKGYPCGRYVHHDRGALVYVDPLKJISKQCKRIQVYHLEB 1543
Qy      1275 VRSALKIFGRCTSELSLYKYVAEAVRDITK 1304
Db      1544 LRESLCDVASNLNCAVFSQDEAVAEVHK 1573

RESULT 11
S38480
nonstructural protein - rubella virus
C:Species: rubella virus
C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 26-Aug-1999
C:Accession: S38480
R:Gillam, S.
submitted to the EMBL Data Library, March 1993
A:Description: Nucleotide sequence of the nonstructural protein genes of rubella virus.
A:Reference number: S38480
A:Accession: S38480
A:Status: preliminary
A:Molecule type: genomic RNA
A:Residues: 1-2115 <Gilt>
A:Cross-references: EMBL:X72293; NID:g410507; PIDN:CAAS1087.1; PID:g410508
C:Superfamily: rubella virus nonstructural polypeptide

Query Match      3.0%; Score 261.5; DB 2; Length 2115;
Best Local Similarity 20.9%; Pred. No. 8e-07;
Matches 330; Conservative 164; Mismatches 594; Indels 491; Gaps 78;

```

```

Qy      163 CAFKRVQIANHSLYDVTLEBELANAFENHGLAMVRAFMHMBEELLNDYNNVNAELGYRFH 222
Db      862 LAADCR-----RLAPCPDGE-AVATPRHG-----CGYH-H 889
Qy      223 VIE--EPMAYVD-CAFOGDLRLHFLDPTINESQERIERLAAAGSYRRAVIFSGDD 279
Db      890 ITHVAPRRPRDPALEBEGALL-----EAYTSIVALAABRRARVACPLGAGV 940
Qy      280 WGDVLYHDFHTWLYLVLRNYPPTFGPSLMT---EYQRN-----GSSIELRTTRAPD 329
Db      941 YGMSAASLSRAALA--ATRAPAE-RVSLIHCHPDRATLTASAVLVGAGLAARVSPPT 997
Qy      330 -----GDRMLAVVPRTSG--LC----- 345
Db      998 EPLASCPAGDPGRPAQASAPATPLGD-ATAPPRGCGGELCRYTRVTNDRAVYVIML 1056
Qy      346 -----RIDNIFPYADSGTEHKTILTSQKVN--MLNFMQTRPEKEV----- 387
Db      1057 ERDQATSMARIRIEVVVY---GPEH---LATHPLNHYSLVLPABVRPRFGCGSDMW 1109
Qy      388 -----DWTVMSPARALRAIVVASEYTBSSMNISPADLVRTVVSILYHIIERR 438
Db      1110 RCRGMQGWMPQVRCYTPSNHAALCRTGVPFRVSTGGEIDP-----NTCMLRAAAYVAQAR 1165
Qy      439 AAVAVKTA-----KDVFGETSFWESLKHVLSGCG-----LRNLKGTDVV 479
Db      1166 ACGAYTSAGCKKAYGRALSEARTHEPFAALSQWMSASHADSPDGTGDPDPLMEITVGC 1225
Qy      480 FTKEV-----VDKRV-HSLGDIICDVRISP-----QVGLPFRV 514
Db      1226 ACSRVWVSSEHARPDLHLVSLHAPNCPMGVLEVAPRPGGAPTHGFVCAVGGGPRRV 1285
Qy      515 PPARVHDESELVLRAGACYNRPVSTPVEBPQGFADLMATAASLEPR-----A 569
Db      1286 -----SDRP-----HLMPLA-----VPLSRGGGTCA 1305
Qy      570 TLQGL-----NTDYKQL--KILENAKTIIDGLTSPVAGLEMYBSPPSGRTGITIA 621
Db      1306 ATDEGLAQAYDDLEVRRLGDAMARAALASIQRPKQPY-NIVWNNMAAGAKTRITLA 1364
Qy      622 ALBAAGKALVYATPRLREAMDRIKPPASATQHVALLRATAGAPATVVDDEC 681
Db      1365 AFTR-----EDLYVCTNLNHLHIOAKLRARDIDIK--AATBRALTYRLAIVRYIYDEA 1419
Qy      682 FMPVLVYVAIVHALSPSGRIVLVDVHQIGFIDQGSANMPLVDVVKCRRTFNQTK 741
Db      1420 FTLAGGYCAFY-ASQTAFLVLCVGDRCQG--PHYANNCRTP-VBDRMPTGRSR---HTW 1472
Qy      742 RCPADVAATTFFQSLYRCCTTSGCVASISHVADYRNSOQT-----L 785
Db      1473 RFP-----DCWAARLRAGLDYDIBGERTGTPACNLMDGRQVLDL 1512
Qy      786 CFTEBESKRGHAGG--AMTVHAGGRFPASVITLYNGSTABOKLAESKHLVIGTRHTN 843
Db      1513 AFSRETVRRLHBAIGRATVRAAGMSGTACIHVGKRDGTVALATLDLAIIVSLTRASD 1572
Qy      844 HLYIRDPGTGDIRQLNHSKAKAVFTDIPAPLEITTV-----KPSSEVQ 886
Db      1573 ALYIH-----ELDDGLRAAGLSAFLDGALAELEKVPAGIDRVVAVBQAPPLPADGIP 1628
Qy      887 RNEVMATIPP-----QSATPHGA-----IHLRKNFGQOP 916
Db      1629 BAQ--DVPPCPPTLEBELVEGRAGHPYADLNVTBGEREVRYMRISRHILNKHRTMP 1685
Qy      917 DCGCVLAKTGYEVFGRAKINVELABDAPKPHAFQEGVQVWVYVNAENKQALQTL 976
Db      1686 GTERVLSA-----VSPCGYRAGEDG-----STLFTA 1711
Qy      977 LSRYTKESADLPHEAKEDVKRMNLSDRMDWTVTBDRADRAVFERQKFTQKGTED 1036
Db      1712 VARQHPR-----PFIQIP--PRVYAGVAQEBRMITYLERIDLVTVYTGMAAR----- 1759
Qy      1037 LLEPDDEPIYR--DIDPLMKTQOKVSRK-----PINT-----GRVGGG 1071

```

Db 1760 --ELDRTYRRRPEIRAGCTAGSLVPAFLKATLKCVDAAAGPRTEDCHAAGAGAG 1817
 QY 1072 IAAHKSILNVLAAIRILEELRTGSR--TVRYNGLPDEEAMLLBAKINQVPHAT--F 1128
 Db 1818 IAAWAKWQVMSPHRAIQIKIMRALRPOFLVAAGHTEPEVDAMQA-----HTTNA 1871
 QY 1129 VSADWTEPD--TAHNTSELPAALLERIGTPAAAVNLFRECGRTLR-----KGL 1179
 Db 1872 IEVDTEFPMNQTLATRDVELEISAL--LGLFCAR-----DYRALRAGSYCTLRRL 1921
 QY 1180 GSVEVGLDLSGAAWTPCRRITIPSAVMTLF--RGVFAA--FKGDDSL-----CGSHYL 1232
 Db 1922 GSTETCERSTSGEPALLHNTTAAACMAAMVKGVRMAGIFGGDDMVIFLPGARNAL 1981
 QY 1233 RFDASLHM--GERYKTKHLKVEVQKIVPIGLLSABOVVLDPRSALKIFGRCYTSEL 1291
 Db 1982 KMTPAEVLFGFHIPIVGHVSTPTSPFCGHV--TAAGLPHDVHQAIKVLGRFPDVL 2038
 QY 1292 YSKYVAVAVDITGNSDARHSLLCMSACVYTAESAAYITDAVVRGRG--DPP-- 1346
 Db 2039 EEOQV--ALDLRLGVAAALPDVAA--NMAVYDSABRLVAILRELTAAVAGRGDLHPAT 2095
 QY 1347 --FEOLRVRAHVQAPDA 1362
 Db 2096 IGALEETQTPYARANLHDA 2114

RESULT 12

S26358
 hypothetical protein, 126k - tobacco mosaic virus
 C/Spectes: tobacco mosaic virus, TMV
 C/Date: 25-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 17-Mar-2000
 C/Accession: S26358
 R/Koh, H.K.; Song, E.K.; Lee, S.Y.; Park, Y.I.; Park, W.M.
 Nucleic Acids Res. 20, 5474, 1992
 A/Title: Nucleotide sequence of cDNA of the tobacco mosaic virus RNA isolated in Korea.
 A/Reference number: S26358; MID:93055219; PMID:1437566
 A/Accession: S26358
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown
 A/Molecule type: RNA
 A/Residues: 1-1116 <KOH>
 A/Cross-references: EMBL:X68110
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, August 1992
 C/Superfamily: Barley stripe mosaic virus alpha-a protein

Query Match 2.9%; Score 253; DB 2; Length 1116;

Best Local Similarity 20.5%; Pred. No. 9.2e-07;
 Matches 234; Conservative 128; Mismatches 387; Indels 394; Gaps 49;

QY 24 QORAVKLDRAFLKALETL--HRLYPLRKGGTLPEYQHPT--LAGHORVAE----- 72
 Db 42 RDRRPVMSKSVISEQTLIATRAYPEFOI--TFYNTQNAVHSLAGLSLELYLMQ 98
 QY 73 -----EVLHNF-----RGRSTVLEIGPSLSALKHGAAPAPADVYGCRTKGRDG 120
 Db 99 IPIYGLTYDIGGFPASHLEKGRAYV-----HCCM-----PMLDV-----RDI 135
 QY 121 SRH-----ITALERSVATGR--PEFKD-----ASILANGIASRTFCVDGVSCAFKS 167
 Db 136 MRHEGGKDSIELVLSRLERGGKTVPNQKAPRVAELPDAVCHNTFQMRHQPMQSG 195
 QY 168 RV-GIANHSLYDVTLSLANAFENHGLMVRATPMHPEELLVMDNVNNAELGYRFVIES 226
 Db 196 RYVAIALHSLYDIPADEFGAALRKQVHTCYAAFHSENLILDSYVNL-----245
 QY 227 PMAVKCAFGQGLRLHFPDLFINSQERRIRIARGSYSR-----RAVPSGDDMG 281
 Db 246 --EINACFSRQD-KLTF--SPASBS-----TLNCHSYSLNKLKVCCTTPASN--R 291
 QY 282 DAYLHDF--HTWL-----AYLLVR-----NYPTPFGSLH 309
 Db 292 EYVMEKELVAVNTVMFCFSRIDTFLYKGVAKHSVDSQFYTAMEDAMHYKXTLAMCNS 351

QY 310 IEVGRHSGSIELRTTRAPQDRMLAVPRTSQGLCRIPNIPYADAGSTBKHTILTSQH 369
 Db 352 ERILEDSSVWYWF--PKQRDMVIVLPDISL-----ETSRRTKEVLVSD 397
 QY 370 KYNMLNFMQTPKBEKLVDMVTLMSPARARLAIIVASBVSSWNISPADIVRTVSLY 429
 Db 398 FYPTLNIHRTQALTYTANVL--SFVESIRSRVLIINGVTANBSMDVDKS--LLOSLSMTF 455
 QY 430 VLIH-----IERRAAVAKTAKDDVPGSTF--MESLK-----HYLSGC 467
 Db 456 YLHTKLAVLKDLLLSKESLSGSKTCQVWDEISLAFGNAFPSVKBRLNKLIVAGDA 515
 QY 468 CGRLNKGDVFTKGVVDKTRVHSLGDIICVRLSPRGVGLPSRVAPVAFHDBEE 527
 Db 516 LBR--VPDLVYTFHDLVTEYKA--SVDMPALDIRKMBETBYM-----YNALSELS 564
 QY 528 VLRBAGCVN-----ERF-----539
 Db 565 VLRSDKPDVDFVSGMCSGLBYDPMTAKVIVAVMSNSGLTLTTERPTEANVALALODQ 624
 QY 540 -----VPSTPVEBPQ-----GPDAD-----LWH-ATA 561
 Db 625 EKASGALVVTSRVYEBBPMKSMARGLQLAGLDHPBSYSRNBIBLSLQFHMATA 684
 QY 562 ASLPEYR-----ATLQGLATDVQQLKITLENALKT-----592
 Db 685 DSLIRKQSSIVYTPIRYQMKNPIDSLVSLSAVSNLVKILDTAIDLETOKGV 744
 QY 593 -----I 593
 Db 745 LDVASRKLIRKTAASHANGVETARKKVALLEYDEQVYTCNMRVAVSSSVYS 804
 QY 594 DGLTLPVNGL-----EMYBGPSSGKTGTLIALBAAGKALVAPTRBLR 640
 Db 805 DNAKRTLRRLRNGSPHVSAAKVLVDCVPCGCKTEILSRVNF--DRLILVPGKQA 862
 QY 641 EAMDR-----IKPSASATQVALALIRRAABGAPRTVVIDCFMPLPVYVAIVHL 695
 Db 863 EMIRRRANSIGIIVATKCNVKTVDSPMNFQKSTQCFRLRIDGLMHTGCNVFLVM 922
 QY 696 SPSSRIVLGVADVHQIFID-----FOGTSANMPLVBDVYQGRRTFNOTKRCPADVATT 751
 Db 923 SLCEIAYVGDVQOQPIYINRVSGFYPAHFLALDEYRZ--TRKT--TLKRPADV--TH 975
 QY 752 FQSLVPG--CTTSGCVASISH-----VAPDYRNSAQTLCTFOEBS--RHAGS 799
 Db 976 YINRRYEGFVMTSSVSKSVSQBMVGAAVINPISKPLHGKITLTFQSDKALLSRGSD 1035
 QY 800 AMTYHAQGRTPASVILHNGSTAEOKLAEKSHLLVGTTRTNHL-----YIRDPTGDI 855
 Db 1036 VHTVHEVQGETYSDVSLVRLPTPVSIIAGDSPHVLVALSRHTCSLKYTVVMDPLVSI 1095
 QY 856 RQL 858
 Db 1096 RDL 1098

RESULT 13

MMNVN

nonstructural polypeptide - rubella virus (strain Theilen)

C/Spectes: nonstructural protein NS1; nonstructural protein NS2; nonstructural protein NS3

C/Date: 30-Sep-1989 #sequence_revision 30-Jun-1992 #text_change 16-Jul-1999

C/Accession: A35320; A29811
 R/Dominguez, G.; Wang, C.Y.; Frey, T.K.

A/Title: Sequence of the genome RNA of rubella virus: evidence for genetic rearrangement

A/Reference number: A35320; MID:90281585; PMID:2353453

A/Accession: A35320
 A/Molecule type: genomic RNA
 A/Residues: 1-2205 <DOM>
 A/Cross-references: GB:M15240; MID:G333971; PIDN:AAA08528.1; PID:G333972

R: Frey, T.K.; Marx, L.D.
Gene 62, 85-99, 1988
A: Title: Sequence of the region coding for vitron proteins C and E2 and the carboxy term
A: Reference number: A29811; MUID: 88226020; PMID: 2836271
A: Accession: A29811
A: Molecule type: genomic RNA
A: Residues: 1737-2205 <PRE>
A: Cross-references: GB: M15240
C: Comment: The cleavage sites of this polypeptide have not been determined.
C: Superfamily: rubella virus nonstructural polypeptide
C: Keywords: nonstructural protein

Query Match 2.8%; Score 248.5; DB 1; Length 2205;
Best Local Similarity 21.4%; Pred. No. 5.1e-06;
Matches 205; Conservative 126; Mismatches 372; Indels 255; Gaps 46;

600 EVRG---LEMEGPGSGKTGLIALALBAAGKALVYPTRELRAMDRIRKPPASATQ 656
1339 PRKGPYNI RYNNMAGAGCTTRILAAFR---EDLYVCTNALHBIQAKLRARDIDIKN 1395
657 HVALAILRATDEGAPFATVVIDECFMPFLVYVAIVHALSPSSRIVLVDVHQIGFIDPQ 716
1396 --AATYERRLTKPLAAYRRIYIDEAFTLGSEYCAEV--ASQTAAYICVGDIDQCG----- 1447
717 GTSANMPLVRDVVQCRRTFNQTRKCPADVATTFPQSLVGCCTTSGCVASISHVAPD 776
1448 -----PHYANNCKTPV---PDRWPEKSRHTW---RFPDCW-----AARLRAGID 1486
777 YRNSQAQT-----LCFTQEEKSRHGAEG--AMTDEAGRTFASVILHY 818
1487 YDIEBERTGTACNMDGRQVDHLAFSEETVRLHENGIRAYTVREHQMSVGTACIHV 1546
819 NGSTAEQYLAEKSHLVGTRHNLHYRD-----PTGDIEROLNSAKAEVFTDIPA 872
1547 GRDGTVALALTRDLAIVSLTRASDALYHLEDESLRAAGLSAFLDAGALAEI--KEVPA 1605
873 PLE-ITTYK-----PSEEVQRNEVMAITP-----QSATPHGA----- 904
1606 GIDRVAAVEAQPRPLRPADGIPDAQ--DVPPFCPTLIELVFGAGHGHYVLDLRNVTG 1662
905 -----IHLKRNFDQDPCGVALAKTGEVFGRAKINVELAPDAPKPRAF 954
1663 EREVYMRISRLLKNTHEMCTERVLAV-----CAVRYRAG 1702
955 QEGVQWVKNVNASNGHQLTLLSRYTKSADLPHEAKEDVKMLNSLDRHMDVTYED 1014
1703 EDG-----STLRTAVARQHPR---PFRQIPR--PRYAGVAGQEMWTYARE 1743
1015 ARDRAVFETQKFTQGGTVEDELRPDDYIR---DIDPLMTQOKVSPK----- 1061
1744 RIDLTDVYVTVGVAAAR-----ELTRVARRYPEIFAGMCTAOSLSVPAFLKATLKCV 1795
1062 -----PINT-----GKVGQIAHSHKSLNFVLAAMIRILEIIRTSGR--TVRYSNGLP 1108
1796 DAALGPRTECHAAQKAGLEIRAKAKEMVQVMSPHFAIQKIMRALRPPFLVAAGHT 1855
1109 DEEAMLLKAKINQVPHAT--FVSADWTEFD--TAHNTSELPAALLERIGTPAAVN 1163
1856 EPEVDAMMOA-----HYTNAI BVDPEFDNMQTLATRDVELISAAL--LGLPCAB-- 1905
1164 LFRKCGRTLR A-----KGLGSEVVDGLDSGAAMPFCRNTISAAVMTLFP--RGYKF 1216
1906 -----DYRALNAGSYCTLRIGSTETGTCERTSGEPATLLHNTYVAMCAMMVPKGVAM 1959
1217 AA-FKGDLSL-----CGSHYLRFPASRLM--GERYKTGHLKVEQKIVPYIGLVSABQ 1269
1960 AGIFGQDMVIFLPEGARSAALAKMTPABVGLGFPNIVKGVSTTPPSFGHG---TAAAG 2016
1270 VVLDEVRSAKIFGSCYTSSELLSKYVEAVRDTIKGMSDARYSHLLCHMSACYNYAPES 1329
2017 LFHDVHQAIKYLCSRFDPDVLVEQOV--ALLDLRLSGVVAALPDTVAA--NAAUYDYSAER 2073
1330 AAYITDAVVRFGRGDFPEQLAVVRAHQVAPD-----AYSGTYPRANVRASCIDHVEPRQA 1385

Db 2074 -----VLAIVR-----ELRAYAGARRPFGHRRARGDSDPLRAAGS-----PRRR 2114
Qy 1386 AAPAGFVATCAKPR--TPBSLLTAKAGVSAATTSNVAATGA-----PPESP 1427
Db 2115 LTPIL-YVGPLILPILTTSSPTVSPHVLGTQLPFGRAFGCPNGFYRHHHGGPRGPR 2171

RESULT 14
NMTMS2
186K protein - cucumber green mottle mosaic virus (strain SH)
C: Species: cucumber green mottle mosaic virus
C: Date: 30-Jun-1993 #sequence_revision 30-Jun-1993 #text_change 08-Apr-1994
C: Accession: J01157
R: Ugaki, M.; Tomiyama, M.; Kakutani, T.; Hidaka, S.; Kiguchi, T.; Nagata, R.; Sato, T.
J. Gen. Virol. 72, 1487-1495, 1991
A: Title: The complete nucleotide sequence of cucumber green mottle mosaic virus (SH strain)
A: Reference number: J01157; MUID: 91311400; PMID: 1856687
A: Accession: J01157
A: Molecule type: genomic RNA
A: Residues: 1-1646 <UGA>
A: Cross-references: GB: D12505
C: Superfamily: cucumber mosaic virus RNA 1 protein

Query Match 2.7%; Score 242.5; DB 1; Length 1646;
Best Local Similarity 17.5%; Pred. No. 7.3e-06;
Matches 282; Conservative 222; Mismatches 559; Indels 547; Gaps 70;

137 PEFKADSLANGLASRTFCVGVGSCAF-----KSRGIANHSLYDVLBELANAFENH 191
162 PTFQIDARRRDSBPACVATGCDVQEGSYDGSGRDNHVAHLSHYDLPYSIGALMRK 221
Qy 192 GLHVRAPFMEPEBILYNDVNA--ELGGRFVIEBPAAVDCAFQGGDLRLHPELDFT 250
222 NVRVCYAAHFSBALIGSPVGNLNSIGAQRV-----DGDVHFLFSEBSTL 269
Db 251 NESQRRTERLAAAGSYRAVIFSGDDMDAYLHDHYTLAYLVANNYPRPESLHI 310
270 HYHSLNIKILVMTY-----PPADDRF--VYIKEP-----MVKRVDTFFPRALRA 314
Qy 311 EVQRHGS-----SIELRITRAPP--GDMYLAIV--PRTSOGLCRIPNIFYVADAG- 358
315 DTHMLHSGVHYSKSKSYPALNTPPIFODATPSVWPEAKVL--IPKELSLFSLGN 372
Qy 359 TEHKTILTSQKVNMLNFMQTRPEKELVDMTVMSPARALRAIVA SEVTESSWNISP 418
373 VKISRMVLADPFTIINHISTYDKAKALVMKNV--QSFVBSIRSRVIVNGSVKSEMVN--P 430
Qy 419 ADVRTVSLVYHIERRAVAVKTKADVPGET-----SPRESLGHV--LQSCC-- 468
431 VDQI--TDISFIFPLVYRK--VOIELMSDKVYIEARGLLRRFADSLSAVEGLDCYVD 487
Qy 469 -----GLRN-----LKGTVVFTYRV 484
488 ALVQGMFDTSSDELKVLPPFPMTFSDYLSGMYTADAKIERBSVSELASGDDLPFK-- 545
Db 485 VDKR-----VHSGLDII CDVRLSPBOVGLPSPVPPAR----- 518
546 IDEIRNNYSGVPEVKEQFCKELNVNPMILGHVIBAFSGKAGVTVTGIGTLPENGA 605
Qy 519 -----VHDBRELEVLRAEGCYNERP----- 539
606 SVALSTSVDTCEMDVTEDMEDIVLMDKSHSYNSPKMARADVKGNNKALVEYKVG 665
Qy 540 -----VPSI-----PVBERPGFDAL-- 556
666 TSMILPATMAEKYRAVLESLGICVAKKPPSKPLREBDRLKSNMPPKYSIDLAKKTI TP 725
Qy 557 -----WHA7PAASLPEYR----- 568
726 VVYGTIERQMKYVIDYLSASLSTGLNLRIVASDMNGTBSHQGTGLDCECKMYL 785
Db 569 -----ATLQAGLNT-----DYKQDK-----ITLBNALKTIDGL 596

```

Db      766 LPAEKHAMAVVLAADDTRRIIFLSYDESGSPIDDKMKRAVCSERYKVSIVISLEYL 845
Qy      597 T-----LSPVRLGEMTEGPGSGKGTGLTIAALEMAGKALYVPTRELRAMRRIKPEPA 652
Db      846 NKEALVDEPVHLLVDGVGCGKTAELIARVN--WKTDLVLPGR-----889
Qy      653 SATOHVALAILRRARA-EGAPPAT-----VVIDECFMEPLVV--AIVHA-----694
Db      890 -----AAAMIRRRACALHKSPVATNDNVNTPDSFVNRRKIKFEDAVYVDEGLMVTGLLN 944
Qy      695 ----LSPSRIVLVGDVHOIGFIDPGSTANMPVADVVKQCRRRFNG-----TKRCP 744
Db      945 FALKISGCKKAFVFGDAKQIPFI-----NRVMFDPYKELRTLLVDVBERRYVTHRCP 997
Qy      745 ADVVATTFPOSLLYPC--TTSGCVASISHV-----APDYRNSOQTLCTFOEBS--793
Db      998 RVV--TSPLNTIYKAAVATTSPPVSHVKAIKVSAGILAPBELTKIKGKITTTQSDKSL 1055
Qy      794 -RHGAGAMTVEAOGRTFASVILHNGSTABOKLAESKSHLVGITRHTNHLIYRDPFG 852
Db      1056 IKSQYNDVNTVHEIQETFEETRAVAVATPTPIGLIARDSPHYLVALTFRAT-----1105
Qy      853 DIEROLNHSKABEVFDIPAPLEITTVKSEVQRN--EVMATIPQSATPHGATHLAK 910
Db      1106 -----KAMVYTVVFDVAVTSIADVERKVDOSILTMFAITVETKOLMONSLVHRN 1155
Qy      911 NFGDOPDGCVALAKTY-----EVFGRA-----KINVELAPDAPTPKPRAF 954
Db      1156 IF-----LPSKGTGFYTDQEFYDRCLPGNSFVLNPDVATMRLRNERLQPCRLT 1207
Qy      955 QEGVQWVKVTNASKQA-----LQTLISR-----YTKRSADLPRLHAKED 995
Db      1208 LSNLDVP--ALIKNEAQNLIPVLRTACERPRIGLLENLVAMIKRAMNPDLAGTVD 1264
Qy      996 VKRM-----LNSLDHMTVYTEDAR-----DR 1018
Db      1265 ITNMSISIVDNFSSFVRDEVLLDHLCVRASSIQSFDMFCQPTSAVGOLANFNFDL 1324
Qy      1019 AVPEQLKFTOR--GGTVEDLLEBDPYRIDIDFLMKTOOKVSPKINTGKVGQGLAHS 1076
Db      1325 PAFDYTMHMTIKOPKSRLLTSIOSEYPALQTVY-----HPKYVN-----AVNG 1368
Qy      1077 KSLNFEAMIRILEEILRTGSRVYVNSGLPDEEBAMLEAKINOVPHATP--VSADWT 1134
Db      1369 PAFKILITKFLMVD-----SSKFFPYTRKKPED-----LQRPESLSSHSDVEILELDS 1419
Qy      1135 EBDTANNTSELLPAALERIGTPAAAVNLFRERCCKRTLRAGLGSVEVDGL-----DS 1190
Db      1420 KYDKSQSDPHPSIEMAIWEKLGDLILAMMW--SMGKRTILDPQAGIKT--LIYYQRKS 1476
Qy      1191 GAAMPFCRNTIFSAV--MLTLFQVKAFAFGDLSLC--GSHYLFQDASRLMGR 1244
Db      1477 GVTTFIGTFTIILACVASMPLDKCFK--ASFQGDLSLIYLPKGLYPRIOQT--ANLVWN 1534
Qy      1245 YTKHLKVEVOKIVPYIG--LLVSAEQVLDPVRSALKIFGRCYSELLYSKYVAVR--1300
Db      1535 FBAKLFR---KKGYGCGKIYIHHANGCIYV--DPLKLIKSLKNGSLVGEHVEFRIS 1589
Qy      1301 --DITKMSDARY-----HSLCHMSACYYNTVPESAAYIYDVARP 1340
Db      1590 LLDVASHSLFNGAVFHLLDDAIHELFPNAGGC-----SFVINCLCKY 1630

```

RESULT 15

S49432

replisase 126K - odontoglossum ringspot virus

C:Species: odontoglossum ringspot virus

C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 19-May-2000

C:Accession: S49432

R:Yyu. K.H.; Park, W.M.

submitted to the EMBL Data Library, October 1994

A:Description: The complete nucleotide sequence of odontoglossum ringspot virus.

```

A:Reference number: S49432
A:Accession: S49432
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1112 <RNU>
A:Cross-references: EMBL:X82130, NID:G558231, PIRN:CA57641.1, PID:G558232
A:Superfamily: Darley scitope mosaic virus alpha-a protein

Query Match      2.7%; Score 242; DB 2; Length 1112;
Best Local Similarity 20.0%; Pred. No. 4,1e-06;
Matches 230; Conservative 129; Mismatches 352; Indels 438; Gaps 56;

Qy      12 YAAADVAAVNLQQAVALDFAPPLKALSTHLRYPLRFKGTLPPOHPILAGHORA 71
Db      66 YAFPEITTYNT--QLAVH-SMAGGIRALB-----LEYRMOQIP-----FGSIT 105
Qy      72 BEVLNPA-----RGSTYLBIGPSLSMLKLGAPNAVAVD--HGCTK--YGTRODSRHT 125
Db      106 YDIAGNFSAHYKGR-----DVHSCMRNLDIRVARRHN 140
Qy      126 ALRSRVATGRPEFPADASLANGLASRTFCVDVGSCAFK-----SRVGLA 172
Db      141 QDDTVSSYVARLERSKGLPVFQGAFFKMSDPPAVCSDKRFGSCYSVDLPKGTAVNG 200
Qy      173 NGLYDVTLEELANFENHGLMVRAPMAMPBELLYMDNVNABIGYRFHVIEBPMAVKD 232
Db      201 LHSIYDIPADBERGALLRKDVHICYAAFIISERLL--ETTSAPL--DEIGATK- 250
Qy      223 CAFQGDRLRHPELDPI-NBS-----QERRIERLA 262
Db      251 --YKSGD-RLSP-----FIONBESTLVNHSYKNVITVCTPTIPASNRFYTHRECFCTRN 303
Qy      263 A-----RGSYSRAVIFSGDDDWGDAYLDFHT-----W--LAVLVNRY 300
Db      304 TWFKCTKVDTYFLRGGYTR-----GEDS-----EGFTYAMDEAMEYKKTILAMNBS 351
Qy      301 PTPPGPSLHIEVORHSGSIELRTAPPGDMLAVPRTSGQLCRIFNIPTYADAG-T 359
Db      352 RTIF-----HRAAVNFR--PKVDWVIVP-----LFGSASTVSGM 386
Qy      360 EHKITLTGQKYNMLNMQTRPEKGLVDMTYLMSPAARLAIYVASVT--ESSWNISP 418
Db      387 KREVMVNDQFYVTLNHIITYODALTYYKNL--SFVSS--IKSRVINGVATARSMEVDK 444
Qy      419 ADL-VRTVSLVYLHIIERRRAVAVAKDVPGETFSFWSLKVLAGSCC-----468
Db      445 SVLQPLSTFTLQTLGAKKQVVLAKFKQKIDDTYNLFWKQISDAVGLPFSIKERLIS 504
Qy      469 -----GLRNKQTD--VFTKRVVDKTRVSHLDIICVRLSPROVGLPSRVPPAR 518
Db      505 GGFVVAEQSLQIKTPDEYITPADKLVMEYK-----ATBELQHLDISKPLER 551
Qy      519 V---FHDBRELEVLREAGCY-----NERVPSPTPV 547
Db      552 AEKTYNALSELVLSDEFDITQRFKNCIREFDIAPDVIAKYVPMKNBELTLPENNPTP 611
Qy      548 E-----FGGPDAD-----LWHTAASLPEYRATLQGL-----575
Db      612 EALSDBLSPLPKDLMRPSLKLSTCAPPPSVK--TLDSLLRKQSGDGRQBSQSVSV 670
Qy      576 -----NTDVNQLK-----ITLENALKTI--DGLTISP 600
Db      671 SDFHLKSVESVKIKMSASAVYTPGLVQGMKMYMDYLSASISATVSNLCKVKLDVGVDP 730
Qy      601 -----VRG-----LEMTBGP-----611
Db      731 SBAEKSGYDVVKGKMIIPKOKCIAMGVADIANGBKVIYLLBMDAGFTICDMRRVAVS 790
Qy      612 -----GSGKGTLLIALEAAGKALYVAP-----TRELRAM--DRRIK 648
Db      791 SDSLIYSDMGKQTLLRCLK--DGEPLVLRMPKVTIVDVGAGGKKEILETVNPPBELLL 848
Qy      649 PPSASATOHVALAILRRATAPGAPPAT-----VVIDECFMEPLV 687

```

```

Db      849 VPGKEACK---MIKXANKSGHVRATKONVTVDSFILMHLKPKTYNKLFIDEGLMLHTG 904
QY      688 YVAIVHALSPSSRIVLVGDVHQIGFIDQTSANMPLVVDVKOC--RRRTFNQTKRCPA 745
Db      905 CVNPLIALSHCREAMVFGDTEQIIPPIN--RVANFPYKHFQHTCLHREVRRLSLRCPA 961
QY      746 DVVATTFPQSLYPG---CTTSGCVASISHVA-----PDYRNSQAQTLCTQOEKSR 794
Db      962 DV--THPMNSKYDKFPLC--TNDVIKSYDAEVVRGKGVFNPKSKPLKGIITFTQSDKAE 1017
QY      795 HGAEG-----AMTVHEAQGRTPASVILHNGSTAEOKLAEKSHLVGITRHTN-- 843
Db      1018 LNERGYEEVSTFGELNIVHEIQGETFEDVSVVRLLPTALBELISKSSPHVLVALTRHTKSF 1077
QY      844 --HLYIRDP 850
Db      1078 KYCCVVLDP 1086

```

Search completed: January 15, 2004, 16:53:04
 Job time : 58.0655 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 13:22:26 ; Search time 23.1935 Seconds
(without alignments)
3454.993 Million cell updates/sec

Title: US-09-991-262-40

Sequence: 1 MYKATDVARYMAADVAYA.....LKIRISTSDSVDPVKISKSA 1704

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_41:*

Precl. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	542	6.1	1691	1	POLN_HEVME Q03495 hepatitis e
2	497	5.6	1693	1	POLN_HEVPA Q03424 hepatitis e
3	488.5	5.5	1693	1	POLN_HEVMT Q04610 hepatitis e
4	479	5.4	1693	1	POLN_HEVMT P23324 hepatitis e
5	327	3.7	1616	1	RRPO_TOMG1 Q9yk46 tomato mosa
6	322.5	3.7	1616	1	RRPO_TOML P03587 tomato mosa
7	321.5	3.6	1616	1	RRPO_TOMK1 Q9q1t8 tomato mosa
8	319.5	3.6	1616	1	RRPO_TOMK2 P88676 tomato mosa
9	301.5	3.4	1612	1	RRPO_ORSVK P88659 odontogloss
10	300	3.4	1612	1	RRPO_ORSVS Q84133 odontogloss
11	298.5	3.4	1597	1	RRPO_PMMV P88657 pepper mild
12	297.5	3.4	1612	1	RRPO_PMMV P23098 pepper mild
13	297.5	3.4	1612	1	RRPO_PMMV Q98745 tobacco mos
14	296.5	3.4	1616	1	RRPO_TMOB P90211 tobacco virus
15	292	3.3	1616	1	RRPO_TMOB Q9yq08 tobacco mos
16	287.5	3.3	1616	1	RRPO_TMOB Q9yq08 tobacco mos
17	284	3.2	1616	1	RRPO_TMOB P30738 tobacco mos
18	280.5	3.2	1616	1	RRPO_TMOB Q88920 turnip vein
19	279.5	3.2	1601	1	RRPO_TMOB P03586 tobacco mos
20	278.5	3.2	1616	1	RRPO_TMOB P18339 tobacco mli
21	265	3.0	1609	1	RRPO_TMOB P18389 rubella vlr
22	248.5	2.8	2205	1	RRPO_SHMV P89202 sunn-hemp m
23	244	2.8	1629	1	RRPO_SHMV P19523 cucumber gr
24	243.5	2.8	1648	1	RRPO_SHMV P03588 bromo mosai
25	228.5	2.6	961	1	VIA_BMV P05080 tobacco rat
26	214.5	2.4	1707	1	VIA_TRVSY P27752 cowpea chlo
27	212	2.4	958	1	VIA_GCMV P18811 equine arte
28	201	2.3	3175	1	RRPO_EAV P03589 alfalfa mos
29	196	2.2	1140	1	YMG6_YEAST P08483 saccharomyc
30	192.5	2.2	1367	1	AMVH_YEAST P08480 saccharomyc
31	185.5	2.1	966	1	VIA_BMV P00020 broad bean
32	182	2.1	725	1	AGH1_YEAST P32323 baccharomyc
33	170.5	1.9			

34	170	1.9	1139	1	VRNA_BSMV P17595 barley scrl
35	169	1.9	1823	1	VIT_ICHON Q91062 ichtyomyzo
36	167.5	1.9	993	1	VIA_CMVNT Q40976 cucumber mo
37	166.5	1.9	993	1	VIA_CMVIX Q66121 cucumber mo
38	165.5	1.9	993	1	VIA_CMVFN P17769 cucumber mo
39	165.5	1.9	993	1	VIA_CMVO P20122 cucumber mo
40	165.5	1.9	2774	1	MAP1_RAT P34926 rectus norv
41	163.5	1.9	993	1	VIA_CMVIT Q83270 cucumber mo
42	163	1.8	630	1	MOCI_MOUSE Q02496 mus musculu
43	162.5	1.8	993	1	VIA_CMVY Q83264 cucumber mo
44	161.5	1.8	1162	1	TCNA_TRYCR P23253 trypanosoma
45	161	1.8	2514	1	POLN_SINDO P27283 sindbis vlr

ALIGNMENTS

RESULT 1
ID POLN_HEVME STANDARD, PRT, 1691 AA.
AC Q03495;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Non-structural polyprotein [contains: RNA-directed RNA polymerase
DE (EC 2.7.7.48) / Helicase].
OS Hepatitis E virus (strain Mexico) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
CX NCBI_TaxId=31766;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93079857; PubMed=1448913;
RA Huang C.C., Nguyen D., Fernandez J., Yun K.Y., Fry K.B.,
RA Bradley D.W., Tam A.W., Reyes G.R.;
RT "Molecular cloning and sequencing of the Mexico isolate of hepatitis
RT E virus (HEV)."
RL Virology 191:550-558 (1992).
RN [2]
RP SEQUENCE OF 965-1691 FROM N.A.
RX MEDLINE=92271462; PubMed=1589964;
RA Fry K.B., Tam A.W., Smith M.M., Kim J.P., Luk K.C., Young L.M.,
RA Pictak M., Feldman R.A., Yun K.Y., Purdy M.A., McCauleyland K.A.,
RA Bradley D.W., Reyes G.R.;
RT "Hepatitis E virus (HEV): strain variation in the nonstructural gene
RT region encoding consensus motifs for an RNA-dependent RNA polymerase
RT and an ATP/GTP binding site."
RL Virus Genes 6:173-185 (1992).
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (RT-NANBH).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL: M74506; AAA45730.1; -
CC PIR: A44212; A44212.
DR MEROPS: C41.001; -
DR InterPro: IPR002589; ALDP.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR007095; RNA_pol_DS_P8.
DR InterPro: IPR007094; RNA_pol_PSVlr.
DR InterPro: IPR002588; V_methyltransf.
DR InterPro: IPR000606; Viral_helicase1.
DR Pfam: PF01661; ALDP; 1.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.

DR Pfam: PF01660; Umechyltransferase; 1.

DR SMART; SM00506; Atp1; 1.
KW Polyprotein; transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding.

FT NP BIND 973 980 ATP (POTENTIAL).
SQ SEQUENCE 1691 AA; 185224 MW; DB3F0B2C913F871B CRC64;

Query Match 6.1%; Score 542; DB 1; Length 1691;

Best Local Similarity 21.8%; Pred. No. 1.1e-24;
Matches 382; Conservative 186; Mismatches 554; Indels 632; Gaps 78;

```

QY 2 YKATDVAVVAADVAAYVAVLQGRVAKDPAPELKLRLHYVP--LRFKGTLPPT 59
DB 6 FIKAPCITTAIEQALAAANALANAVVVRPFLSHQOVELINLMQPROLVRRPEVF--W 63
QY 60 QHPILAGHGVABEVLHNFARGS--TVLEIGSLHSLKLGAPNAPVADYHGTCTKGT- 117
DB 64 NHPI-----QVYHNELEQYCRARSGRCLEIGHAPRS--INDPNV-----LHRCFLHPVG 112
QY 118 RQGSRRITALE-----SRVATGRPEKADASLANGLASRTFCVDDGVGSCAFKSRVG 170
DB 113 RVORVRYTAPTRGPANCRSSALRGLP-----ADRTYCFDGFAGCRFAETG 160
QY 171 IANHSLYDVTLLELANAFENHGLHWRAFEMPEBELLYMDNVYNAELGYRPHVIBPMAY 230
DB 161 VALYSLHDLQPADVAEMARHGKTRLYAAPHLEPEVLL-----PPGTY 203
QY 231 KDCAFQGGDLRLHPELDELINESQERRIELLARGSYSRRAVIFSGDDMDGAYLHD--- 287
DB 204 RISSY---LLIH-----DGKRAVVYVEGSDTSAG--YHNDVAT 235
QY 288 FHTWL--AYLLVRYPTFPFGSLHIEVQRHGSSELRITRAP-----PGRMLAVVP 338
DB 236 LRTWRRTTKVYGEHP-----LVIERVKGICHFVLLITAAPSPSPMPVYPYRSTEVY 289
QY 339 RFSQGLCRIPNIFYVADAGTEHKTLITSOHKVMMLNFQGTPEKELVDMVTLMSFARA 398
DB 290 RSLFGGSGPSLFTPLCAVAKSTHAYPT--HMDRLMLFGATLDDQAF--CSRLMTYLLG 346
QY 399 RLRAIVVASEVTSSWNISPADLVRTVSLY--VLH-----I 433
DB 347 ISYKTVGLVLANEGNATEDALTAIVTAAYLTICQRYLRTQALSKGRMLELHAQKF 406
QY 434 IERRRAVAVKTAADV-----FGETSFWES-----LKVY 463
DB 407 ISRLYSLWLFPEKSGRDYIPGRQLQFYACRRMLSAGFHLDPRTLVPDESVPSCRTTIRI 466
QY 464 LGS--CCGLRNL-----KGTDV-----VFTKRVVD-- 486
DB 467 AGKFCFCFMKVLQGECCSLQPAEGLAGDQCHNEAVEGSDVDTAEPATLDTITGSIYVGR 526
QY 487 -KYVHSLGDIICDV-----RLS----- 503
DB 527 SLQTVQALDLPADLVARARLSATVVTETSGRLDQCTWIGKTFITTFVDCARLEVNG 586
QY 504 PEGVGF-----LPSV-----PPARVP 520
DB 587 PEGNLNLSFDSQCSMAAGPCLTYAAVDGLLEVHSTAGLESRVPPPGNAPTAAPSEVT 646
QY 521 -----HDRB----- 525
DB 647 AFCSALYRNKRGQROSVIGSLMLHPEGLGLFPFSPGHMKSANPFCGESITLYRTWS 706
QY 526 -----LEVLRKAGCTNBRPVSTPV----- 546
DB 707 TITDTPLTVGLISGHLDAAPHSGRPATATGPAVGSDDSPDPLPDVDTGSRSPGARPA 766
QY 547 -----BEPOG 551
DB 767 GPNPNGVPPORLLHTTYPDGAKIYVGSIFSESECTWLVNASNAGHRPGGGLCHAFQRYPDS 826
QY 552 FDA-----DLWHTATASLPEYR-----ATLQAG----- 574

```

```

DB 827 PDATKFWBDDGLAATLTLPRIIHAVA---PDYRLHNPKULEAVRETCARGTAAVPL 883
QY 575 LNTDVKOLKITV-----ENALKTIDGLTSPVR----- 602
DB 884 LDAGIQVAVPVSISPDAMERNRPFDELYLTLELAAPFSSNRPQPTLNTTEDTABAANLA 943
QY 603 -----GLEWE-----GPPGSGKTGTTLTAALBAAGKLYVAPTRBL 639
DB 944 LELDSGSEVRACAGCKVBPVVRVYQTPAGVPSKSKSVQA-----DVDVVPTRBL 998
QY 640 REAMDRRIKPPESASATOHVALAIL--RRATAGAPFATVVIDECMPPLVVAIVHALSP 697
DB 999 RYAMRR--GRAATPHTAARVTSGR-----VIDEASLP--PHILLH--MQR 1043
QY 698 SSRIVLVGDVHOIGTIDPQGSANPPLVADVVKQCRRTFNTQKCRADVATTEPQSLY 757
DB 1044 AASVHLGDPNQIPAIIDEHTL--PAIRP--ELVPTSMVHTHCAPADV--CELVRGAY 1098
QY 758 PGCTTSGCVASISHVADYRNSQOTLCFQEBKSRBAGAMTVHBAQGRTPASVILH 817
DB 1099 PFIQTTSKYVNSLFWGEP--AVGQKLVFTQAAAH--PSTIVHBAQATPTTTTII 1152
QY 818 YNGSTAOKTLAEKS--HLVGTTRTNHLYTRDPTGDIRQNLNS-----AKAEV 866
DB 1153 ---ATRADRGLQSSRAHAIVALTRHTKCVILDSRG--LAREVGISDAIVNPFISGSEV 1208
QY 867 FTIDIPAPLEITTVKSESVQRN--EYMATTPQSATPHGAIHLIRKQFGDPQCCVVALK 925
DB 1209 GHQRP-----SVIRGNPDNRVADVLAAPFPSCQI--SAFHQLAELGHRP--APVAAVVP 1259
QY 926 TGYEVFGRAKINVELAEPDAPKPHAFQ--BGVQVWVYVKNASNGHQLQTLISRYTKS 984
DB 1260 PCPELBOGLLYPQBLASCDVV---TFLTDIYHCRMAAPSORAVLSTLVGRIGRKT 1315
QY 985 A--DLPLHEAKEDVAKMNLSDRNDWTVTEBDARDAVETOLKTFORSGTVEDLLEPD- 1041
DB 1316 RLVDAGHTDVRSALRPIITLGR-----VT--ATTCLEFELVBAWEKQDGSVALEIDL 1368
QY 1042 -DPYTRIDPLMKTOOK--VSPRPINTGKYGQIAHNSLSNLFVLAAMIRILEE--ILRTGS 1098
DB 1369 CSRDSRITFPQDKCKFTTGSTIAHGKQGGIIPMSKTFCALFGPMFAIKALISLPL 1428
QY 1099 RTVRYSNGLPDEBEMMLEAKINOVPHATFVSADMTBPTANNSTSELFALLRIRIGP 1158
DB 1429 QAVFTGDAYDS---VPSAAVAGASHAVFENDSBPSTONNSLSLGBALMBSCGMP 1484
QY 1159 AAAYMLFRE-----RCGKRTLRA--KGLGSEVDGILLDSGAAMTPCRNTIFSAAYML 1208
DB 1485 QMLVRLYHAVRSAMTLQAPKESLRGFMKKHSG--EPGSL-----W---NTVMMAIIA 1533
QY 1209 TL--FRGVKPAAPKGGDS--LLCGSHYLRPDASRLHMGERTYKHLKVEYOKIVPIYGLLY 1265
DB 1534 HCYBEPDQLQVAAFKGDDSVLCSBYRQSPGASLIAGCLK---LKADRPIGLYAGVVV 1590
QY 1266 SAEQVVL--DPVRS 1278
DB 1591 APGLGALPDVVRPA 1604

```

RESULT 2

POIN_HRPVA
ID POIN_HRPVA STANDARD; PRT; 1693 AA.
AC P33424;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Non-structural polyprotein [Contains: RNA-directed RNA polymerase
(BC 2.7.7.48); Helicase]
OS Hepatitis B virus (strain Pakistan) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis B-like viruses.
OX NCBI_TaxId=33774;
RN [1]

RP SEQUENCE FROM N.A.
 RX MEDLINE=92115700; PubMed=1731327;
 RA Tsarev S.A., Emerson S.U., Reyes G.R., Tsareva T.S., Legters L.J.,
 RA Malik I.A., Iqbal M., Purcell R.H.;
 RT "Characterization of a prototype strain of hepatitis E virus";
 CC Proc. Natl. Acad. Sci. U.S.A. 89:559-563(1992).
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N)
 CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
 CC ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>
 CC or send an email to license@ebi.ebi.ch).

CC EMBL, M80581; AAA45725.1; -.
 DR PIR; A38196; A38196.
 DR MEROPS; C41.001; -.
 DR InterPro; IPR002589; A1p.
 DR InterPro; IPR001788; RNA_dep_RNAPol2.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR InterPro; IPR002588; V_methyltransfer.
 DR InterPro; IPR006065; Viral_helicase1.
 DR Pfam; PF01661; A1p; 1.
 DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
 DR Pfam; PF01443; Viral_helicase1; 1.
 DR Pfam; PF01660; Vmethyltransfer; 1.
 DR SMART; SM00506; A1p; 1.
 DR Polyprotein; Transferrase; RNA-directed RNA polymerase; Helicase;
 KW ATP-binding.
 KM ATP-binding.
 FT BIND 975 982 ATP (POTENTIAL).
 SQ SEQUENCE 1693 AA; 185149 MW; 5A0F03B1F1B99E8 CRC64;

Query Match 5.6%; Score 497; DB 1; Length 1693;
 Best Local Similarity 21.2%; Pred. No. 6.2e-22;
 Matches 371; Conservative 185; Mismatches 576; Indels 614; Gaps 72;

QY 2 YAKATDVAARYAADVAVAVVLOQRAVKLDFAPRLKLETLHLRYF--LRFKSTLPPT 59
 DB 6 FIAKPGITTAIEQALAAANSALANAVVRPFLSHQIEILINLMOPROLVFEVE--W 63
 QY 60 QHPLAGHQRVAEEVLNPFARGRS--TVLEIGPSLSALKLHGAPNAADVAGC--TKYGT 117
 DB 64 NHPI-----QVINELELYCRARGRCLEIGAHPRS--INDPNV---VHRGFLPAG 112
 QY 118 RDGSRHITALE-----SRSVATGRPEPKADASLANGLASRTFCVDVGSSCAFKSRVG 170
 DB 113 RDVQRWTTAETRCGPANCRSSALRGLP-----AADRTYCFDGFSGCNPATGT 160
 QY 171 IAHSLYDVLEELANAFENHGLMVAFAFMPEBELLYMDNVNVAELGYRPHVIEEMAV 230
 DB 161 IALYSLHDMSPSDVAELMFRHGMTRLYAALHLPREVL-----PGTY 203
 QY 231 KDCAFQGGDLRLHPELDFINESQERIERLARAGSYRRRAVIFSGDDMDAYLND--- 287
 DB 204 RTASY---LLIH-----DGRV-----VYIEGDTAG--YNHDSN 235
 QY 288 FHTWLAVLVNRYPTPGFSLHIEVORHSSIELRTTRP-----PGDRLAVVPR 339
 DB 236 LRGW-----IRTKVTGDHPLVIERVRAIGCHFVLTLTAAPESPMPVFPYRSTEVYVR 290
 QY 340 TSGCLCPINIFUYADASGTEHKITLSQHYVMMLMFMQTRPEKELVDMTVMSPAR 399
 DB 291 STFGPGSTPLF--PTSCSTKSTFHAVPAHIMWRLMFGATLDQAF--CSRLMTYLRGI 347
 QY 400 LRAIVASEVTESMNISPADLVRTVSLV-----LHIERRAVAV 443
 DB 348 SYKVTGTLVANEGMNASBDALTAVITAAYLITGHORYLRTQALSKMRRLERERHAKFI 407

QY 444 -----KTAD-----DVGRTSFBSL-----KRV 463
 DB 408 TRLYSLMFKSGRDYIPGRQLFPAQCRRRLSAGFLDPRVLVFDSEAPCHCRTAIRKAV 467
 QY 464 LGSCC-----GLRNLKGTVDVPTKR 483
 DB 468 SKFCSPKMLGQSTCETCELPABGVGDQGHNDENAEBSVDPAASASISDLSGVVGVTA 527
 QY 484 VDKTRVHSL-----GDIICDVL----- 502
 DB 528 LQFLYQALDIPASIVARAGLTATVTKVSYVDAGIDCETLLGNKTFRTSPYDAVLETPNGP 587
 QY 503 ----- 502
 DB 588 ERHNLSPDASQSTMAAGPFSLTAAASAGLEVRYAAGLDRAVFPAGVPRGAPGEVTA 647
 QY 503 ----- 502
 DB 648 PCSALYFRNREAOQLSLTGKFWHPBGLGPAPRPSGHVWESANPFCGSESTLYTRTWS 707
 QY 503 -----SPEQ--VGF-----LPSRV-----PPA----- 517
 DB 708 VDAVSPDAQDLPFTSPSIPSPAATPTPAAPLPAPAPDPSPTLSAPARGAPAGATARA 767
 QY 518 ----- 517
 DB 768 PAITQOTARRHRLFTYPDGSKVPAGSLFESTCTWLVNANVDHRRPGGLCHAFFORYPA 827
 QY 518 -----RVHDEBELVLRBAGCYN-----ERVPVS 542
 DB 828 SPDAASFWMDGAAATLTTPRIIHAVAPDRIEMHFKRLBAAYRETSRLGTAAYFLG 887
 QY 543 TPVYSEPOGPDADLMH-----ATAALPEYRA-----TLQGLNTDVQK--LK 583
 DB 888 TGAYQVIGSPFAMERNHRPGDELYLPBLAAMPFANRPTCPLT--ITBEDVARTANLA 945
 QY 584 ITLENAL--KTIDGLTSLFVRGLMEYB--GPPGSKITLTAALBAAGKALYAPTR 637
 DB 946 IRLDSADVGRACAGCVTP--GVVOYQFTAGVPGSKSRSTIOA-----DVDVVVPT 998
 QY 638 ELREAMDRIKPPSASATQVALAL--RATABGAPFATVVDCEMPLVVAIVHAL 695
 DB 999 ELRNAMERR--GRAATPTPTAARVTOGRN-----VVIDEAPSLP--PHLLHLH-M 1043
 QY 696 SPESRIVLVDVHIGEIFDQGTSAAMPFLVADVKKCRRTTFNQTKCPADVAVATTFQOS 755
 DB 1044 QRAATVHLGDPMQIPADIDEPHAGL--VPAIRPDL--APTSMWYTHHCPADV--CELRIG 1098
 QY 756 LYRQCTTSGCVASISHVADYRNSAQTLCTQOBSKSRGAGAMTVHBAQGTFASVI 815
 DB 1099 AYPMIQTTSLVLSLFWGEP--AVGOQKLVFTQAAKLAN--PQSVVHBAQGTAYTETT 1152
 QY 816 LHYNSFABQKLABKS--HLVGTITRTHNLVLRDPTGDIIRQNLNSAKAEVPTDIPAR 873
 DB 1153 IT--ATADKGLIQSSRAHAIYALNHTKCYIIDRG--LMEVGIS-DATVANNFLAG 1207
 QY 874 LBITTVRSEEVQRN-----EVMATIPQSATPHGAILHRLKNGPDQDCSCVALATGX 928
 DB 1208 GBIGHQRPSVIPIRGNPNANDTLAAPPSCQI--SAHQDLABELGHR--APVAVALPCCR 1264
 QY 929 EYFGRAKIVELAEADPATPEKPRAPQ--EGVQYVKTVAASNKQALQTLSSRYTKSADL 987
 DB 1265 ELEGGKLYLPOBELTTCSSVY--TFELTLIVICRMAAPQQRKAVLSTLVRGYGRRT-- 1317
 QY 988 PLHEAK--EDYGRMLNSLDHRMDMTVEDADRAVFEQLKFTQRGGVVEDLLEPD--DPY 1044
 DB 1318 KLTNASHSDVR--DSLARTIPAIQVQVTTCELYELVBAWBEKGODGSANLELDLCSR 1374
 QY 1045 IRDIDPLMKTQOK--VSEKPIITNGKVGIGIAHASKSLNFWLAAMIRILER--ILRTGSRTRV 1102
 DB 1375 VSRITTFQKCNKFTTGETIAGHKVGGISAWSKTFCALRGPMFRAIKKAILMLPGGVF 1434

QY 1103 YSNGLPDEE-EAMLEAKINOVPHATFVSADTEPDTANNTSELLFALLERIGTPAA 1161
 DB 1435 YGADAPDDTVFSAVAARASV-----FENDSEPDSTQNNPSLGLCECAIMECCMPWL 1489
 QY 1162 VULFERCGKRTLRAKGLGSEVDCG-----DSCGAAMPCCNTIFSAVMVLT--FRGV 1214
 DB 1490 IRLY-----HLIRSAMILQAKESLRGKSGSEPGTLLMNTWMAVITHCHDFRDL 1543
 QY 1215 KFAAFKGDST-LCGSHYLRFPDASRLHNGERYKTKLKYEVQKIVPYIGLVSABQVYL- 1272
 DB 1544 QVAAFKGDDSYLCEBYRSGPAAVLIAGCGK--LKVDPRPIGLYAGVVAAPGLALP 1600
 QY 1273 DPVRS 1278
 DB 1601 DVVRA 1606

RESULT 3
 ID POLN HEVNY STANDARD; PRT; 1693 AA.
 AC Q04610;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Non-structural polyprotein (Contains: RNA-directed RNA polymerase
 (EC 2.7.7.48); Helicase).
 OS Hepatitis E virus (strain Myanmar) (HEV).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage;
 OC Hepatitis E-like viruses.
 OX NCBI_Taxid=31769;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=9327573; PubMed=8470371;
 RA Aye T.T., Uchida T., Ma M.Z., Iida F., Shikata T., Ichikawa M.,
 RA Rikihisa T., Wain K.;
 RT "Sequence and gene structure of the hepatitis E virus isolated from
 MYanmar";
 RL Virus Genes 7:95-109(1993).
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 [RNA] (N).
 CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
 ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NA/NBH).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.1sb-sib.ch/announce/
 or send an email to license@1sb-sib.ch).
 CC -----
 CC EMBL: D10330; BAA01172.1; -
 DR MEROPS: C41.001; -
 DR InterPro: IPR002589; A1pp.
 DR InterPro: IPR001788; RNA_dep_RNApol2.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR InterPro: IPR002588; V_methyltransfer.
 DR InterPro: IPR006066; Viral_helicase1.
 DR Pfam: PF01661; A1pp; 1.
 DR Pfam: PF00978; RNA_dep_RNApol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransfer; 1.
 DR SMART: SM00506; A1pp; 1.
 DR PolyProtein: Transferase; RNA-directed RNA polymerase; Helicase;
 KM ATP-binding.
 FT NP_BIND 975 982 ATP (POTENTIAL).
 SQ SEQUENCE 1693 AA; 185215 MW; AAB4C9140V/E21EA CRC64;

Query Match 5.5%; Score 488.5; DB 1; Length 1693;
 Best Local Similarity 21.2%; Pred. No. 26-21;
 Matches 372; Conservative 190; Mismatches 534; Indels 659; Gaps 78;

QY 13 AAADVAANVLCQAVKUDPAPPL--KALFTLRLYV--LRFGKGTLPPTQHPILAGH 68
 DB 22 AAANSALANAVVR-----PFLSHQOIKELINIMQROGVFREVE--WNHPL-----Q 68
 QY 69 RVAEVLNFAFGRS--TVLEIGPSLHSLAKLHGAPNAVADYHCTKYGK--RDGSRHITA 126
 DB 69 RVINHELELYCRAASGRCLTIGAPRS--INDPNV-----VHGFLLPVGRDQVWRTA 121
 QY 127 LE-----SSVATGRDEPADASLNLGASRTFCVDGVGSCAFKRGVIGANHSYDV 179
 DB 122 PTRGPANCRSALRGLEP-----AVDRTYCIDFGSGCNFPAETGIALYSLHDM 169
 QY 180 TLBELANFENHGLMVAAPMMPBELLYMNVNVAELGYRFEVIEEMAVKDCAFQGD 239
 DB 170 SPSDVAEMFRHGMRTLYAALHLPREVLL-----PGTYRTASY----- 208
 QY 240 LRLHPEPLDFINESQERIERLAARGSYRBAVIFSGDDMDAYLHD--PHTLYAVIL 296
 DB 209 LLIH-----DGRV-----VVTYEGDTAG--YNHVSNLRSH----- 239
 QY 297 VKNVPTPGSLHIEVQRHSGSIELRTTRAP-----PDRMLAVPRTSOGLCRIP 348
 DB 240 IRTKVTGDHPLIERVVAIGCHFYLLTAAPBSPMPYVPYPRSTBYVNSIFPGGTP 299
 QY 349 NIFYADASGTEHKTILTSQKVMNLNFMQTRPEKELVDMTVLMSPARALIVASE 408
 DB 300 SLF--PTSCSTKSTFHAVPAHIMDLFGATLDQAC--CSRMTYIRGISTYKVTGTL 356
 QY 409 VTBSMNISPADLVRTVSVLV-----LHIERRAAVV----- 443
 DB 357 VANBGMASBDALTVITVAIYLITCHORYLTKQALSKRMRLERHAKFTIPLYSWLPE 416
 QY 444 KTAQDVFG--ETSPWESLKHVLS-----CC----- 468
 DB 417 KSGRDYIPGRQLFAYQGRWLASGFLDPRVLVDSBAPCRCTAIRKALSKFCFMKM 476
 QY 469 -----GLRNLKGTIVYTKKVVQKRYHS 492
 DB 477 LGQECTCELPAGVGVDGQHDNEAYESGDVDPASALSDISGYSVVGTAALQPLLYALD 536
 QY 493 L-----GDITCDVRL-----SPRQV----- 508
 DB 537 LPALVAAAGRLTATVKVSYQVDGRIDCETLGNKTFRSPFDGAVLEANGERRVYLSFDA 596
 QY 509 -----FLPS--RVPPARY----- 519
 DB 597 SOSTAAGPSLTVAASAAGLEVRYAAGLDHRAVFAVGVSPRASAGETVAFCSALYRN 656
 QY 520 -----FHD-----R 523
 DB 657 RBAQRHSITGLMFWFBEGLIGLFAFSPDGHVWESANPFGSELTLYTRTWSEVDVSSPAR 716
 QY 524 BELVLRBAGCYNER-----PVSTPVEPFG----- 551
 DB 717 PDLGLMSBSLPSRAATPTTLAVLPPAPDPPPPSAPALDEBPASGATAGAPATHQNR 776
 QY 552 -----FDADLMHAT-----AASL-- 564
 DB 777 HRRLLFTYPDSKVPAGSLFBSCTWLVNASVNDHRRPGGLCHAFYQRYPASFASTFM 836
 QY 565 -----PYR-----ATLQAG-----LNTDVQ----- 581
 DB 837 RDGAAYTLTPRPIIHAVAPDRLRBNPKRLAAARCTCSRLGTAAYSLGLGIVPPIG 896
 QY 582 -----PYR-----ATLQAG-----LNTDVQ----- 581
 DB 897 PSFDWERNRHRGDELILPELAAMFBAAPRPTTLTTBVDYARTANLAIBDSATDVGK 956
 QY 592 TIDGTLSPVRLGYEY--GPPSGKGTGLIALBAAGKALVYAPTRBELBANDRIK 648
 DB 957 ACAGCRVTP--GVVGYGTAGVPGSGKRSITQA-----DVDVVVVPTRELANVARR-- 1007
 QY 649 PPSASATQHVALLI--RRATREGAPFATVIVDECPMPPLVVAIVHALSPSRIVLVGD 706

```

Db 1008 -GFAAFPHTAARTQGR-----VIDEAPSLP-PHLLH-MQAAATVHLGD 1054
Qy 707 VHQIGFIDFOGTSANMLPVDVVKQRRRTNQTCKCPADVATTPQSLVPGCTTSGC 766
Db 1055 PNOIPALIDFENAGL-VPAIRPDLGP--TSMWVTHRCPADV--CELIRGAVPMIQTTSRV 1109
Qy 767 VASISHVAPDRNSQAOTLCTQSEKSHGAGMTYHEAGCRFPASVILHYNSTABOK 826
Db 1110 LRSLEWGP-----AVGQVLFVTOAKKAN--PGSVTHAGAGATYTTTII--ATADAR 1160
Qy 827 LLAERK--HLIVGTRTHNLHYIRDPGDIERQNSHAKAFVFDIPAPLEITTKPSEB 884
Db 1161 GLIOSSRAHAIVALTHTREKVIIDAPG-LIREVGIS-DALVNNPFLAGSIGHQR8VY 1218
Qy 885 VORN-----EVMATIPQSATPHGAHILARNFGDDPGCGVALAKTGYEVEGRKXNV 939
Db 1219 PRGNPDANVDTLAFPPSCQT--SAFHQLABELGHRP-----V 1254
Qy 940 ELAEPDATTPKRPAFGGVGV-----KTN-----ASNKHA-LQTLIS 978
Db 1255 PVA---AVLPCEPELEGGLVLPQGLTACDSVTFELTDIVHCRMAMPNORKAVLSTLVG 1311
Qy 979 RYTKRSADLPLEAK-EDVKKMLNSLDRHMDWTEDARDAVPEOTLKTORGATVEDL 1037
Db 1312 RYGRRT---KLYNASHSDVR---DSLARFIPAIPVOVTTCELYELVEMVEKQDGSAY 1365
Qy 1038 LEPD--DPYIRIDIFLMTKTOOK-VSPKINTKVGOGIAHKSANFLVLAIRILEE-1 1093
Db 1366 LEIDLCDRVDVSRITFPQDCNKFPTGERTIAHGKGGISAMSKTFCALFGWFPALERAI 1425
Qy 1094 LRCSRVVRSNGLPDEE-EMMLEAKINQVPHATFPSADTEEDTANNTSELFPALL 1152
Db 1426 LALLPQGVFGDADVDVPSAAVAASV-----FENPSEBDSQNNFSLGECALIM 1480
Qy 1153 ERISTPAAVNLFRERCGKRTLRAGLGSVEVDG-----LDGAAVPCNTTIFSAVM 1207
Db 1481 BECCMPQWLRLY-----HLIRSAMLIQAPKESIRGKWKKGSEPTGLMNTWMAVY 1534
Qy 1208 LTL--FRGVKPAAPKGDST-LCGSHYLRPDASRLHNGERYKTKLKYEVOKIPIYGLT 1264
Db 1535 THCYDFRDFQVAAPKGDSTLYLCEBYSRPSGAAVLIAGCGLK---LAKDPFPIGLYGVV 1591
Qy 1265 VSAEQVVL-DPVRSA 1278
Db 1592 VAPGLALPDVVRFA 1606

```

RESULT 4

```

POLN_HEVBU STANDARD, PRT; 1693 AA.
AC P29324,
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Non-structural polyprotein [Contains: RNA-directed RNA polymerase
  (EC 2.7.7.48); Helicase].
OS Hepatitis E virus (strain Burma) (HEV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=3167;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92024067; PubMed=1926770;
RA Tam A.W., Smith M.M., Guerra M.E., Huang C.-C., Bradley D.W.,
  Fry K.E., Reyes G.R.;
RT "Hepatitis E virus (HEV): molecular cloning and sequencing of the
  full-length viral genome."
RL Virology 185:120-131(1991).
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
  (RNA) (N).
CC -1- MISCELLANEOUS: HEPATITIS E VIRUS IS THE MAJOR CAUSATIVE AGENT OF
  ENTERICALLY TRANSMITTED NON-A, NON-B HEPATITIS (ET-NANBH).

```

```

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ch/announce/
CC or send an email to license@ebi.ebi.ch).
CC -----
DR EMBL; M73218; AAA45734.1; -.
DR PIR; A40778; MNMWH.
DR MEROPS; C41.001; -.
DR InterPro; IPR002589; A1pp.
DR InterPro; IPR001788; RNA_dep_RNApol2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR002588; V_methyltransf.
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF01661; A1pp; 1.
DR Pfam; PF00978; RNA_dep_RNApol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01660; Vmethyltransf; 1.
DR SMART; SM00506; A1pp; 1.
KW Polypeptide; Transferase; RNA-directed RNA polymerase; Helicase;
KW ATP-binding.
FT NP_BIND 975 982 ATP (POTENTIAL).
SQ SEQUENCE 1693 AA; 185191 MW; 2F355846B9ED219B CRC64;
Query Match 5.4%; Score 479; DB 1; Length 1693;
Best Local Similarity 21.2%; Pred. No. 7.0e-21;
Matches 369; Conservative 188; Mismatches 577; Indels 610; Gaps 73;
Qy 2 YAKATDAVAAVAAADVAVAVLQRAVKDLPAPKALKTLHLRYLTP-LRFGKGLTPT 59
Db 6 FIKAPGITTIAIQALAAANSAALANAVVVPFLSHQOIRILINMQPRLQVFRPEVP-W 63
Qy 60 QHPLAGHVAEVLNLPARGS-TVLKIGPSLSHLKLGAPNAVADYHGC-TKCT 117
Db 64 NHPL---QVHNELLYCRASGRCLTGAPRS--INDNPNV---VHRCFLPVG 112
Qy 118 RDGSRHITALE-----SRVATGRPEFADASILANGIASPFYVDGVGSCAFKSRVG 170
Db 113 RDVGWWTATPTRGAAACRSALRGLP-----AADRTYCLDGSGCNPAPATG 160
Qy 171 IANSLVDVTLBELIANFENHGLMVAFAMGMBELLVMDNVNVAELGYRHYIEBPVAV 230
Db 161 IALYSLHDMSPSDVAEAMFRHGRTRLYAALHLPREVL-----PGTY 203
Qy 231 KDCAPQGGDLRIHPELDFINESGERIIEELARAGSISRAVYIFSGDDMGDAYLHD-- 287
Db 204 RTASY---LILH-----DGRV-----VVTYEGDTSAG--YNHVSN 235
Qy 288 FHTVAVLVANVYTPPGFSILHTEVORRHGSIELRITRAP-----PGDRMLAVVPR 339
Db 236 LRSH-----IRTTVNGDHLVIERVAIGHFULLTAAPBSPMYVYPRSTBYVR 290
Qy 340 TSGQLCRIPNIFYAASGTEHKITLTSQHKVNNLMFMQTRPEKELVDMTVLMSPAR 399
Db 291 SIFPGGTSPISF--PTSCSTKSTFHAIVAIWRLMLFGATLDQARC-CRLMTYLRGI 347
Qy 400 LRAIVASVETBSGWNISPADVARTVSVLY-----LHIERRAVAV 443
Db 348 SYKVTGTLVANGWNSBDALNAVIAALTYTICORYLRTQALSKGRRLREHAKFI 407
Qy 444 -----KXAKDVVG-ETSPWESLKHVLS----- 466
Db 408 TRLYSWLPEKSGRDYITGRQLSEFYAQCRRNLSSAGFLHDPVLVVDSEAPCHCAIKAL 467
Qy 467 ---CC-----GRNLTGTDVFTKR 483
Db 468 SKFCFCFKMLGQECTFLQPAEAGVGDQHDNEAVSGSDVDPASALSDSGSVVVPOTA 527
Qy 484 VDKYRHSL-----GDICIVRL-----SP 504

```

Dh	528	LQPLVQLADLPABIVABAGRLTATVVKVSQVVDGRIDCEITLGNKTFRTSPFVDGAVLETNGP	587
Qy	505	EQ-----	515
Dh	588	ERNHLSFDAOSTMAAGPFSLTVAASAGLEVRVYAAGLDHRAVAFAPGVSRSAAGEVTA	647
Qy	516	-----PARVHD-----	522
Dh	648	FCSALYFPNREAOHSLIGNLHFHBEGLGLFAPSPFGHVESANPFCBSSTLYTRTWS	707
Qy	523	-----REBLEVLREAG-----	551
Dh	708	VDAVSSPARPDLGWMSBSPISRAATPTTLAFLPPRAPDPPSPAPRLAPBAGATGA	767
Qy	552	-----PDADIMHAT-----	560
Dh	768	PAITHQTAHRRLFTYPDGSKVPAAGSLFBSSTCTWLVAASVNDHRPGGLCHAFYGRYA	827
Qy	561	-----AASL-----PEIR-----ATLOAG-----LN	576
Dh	828	SFDAASVWMDGAAAYLTTPRPIIHAVADYRLHNPRLEAAYRETCSRIGTAAYPILG	887
Qy	577	TDVQK-----	585
Dh	888	TGIYQVPIGSPFDMERNHRPODELYLBELAAWFEARPFRTLTIEDVARTANLAI	947
Qy	586	LENAL--KITDGLTSPVRLMEYE--GPPGSGKTGLIALEAAGKALVYAPREL	639
Dh	948	LDSATDVGRACAGCRVTP--GVVQYQFPAVGVSGKSRSTIOA-----DVDVVVFPREL	1000
Qy	640	READRRKRPSPASASTOVATAIL--RAATLEGAPFATVVIDECMFPLVYVAIVHALSP	697
Dh	1001	RNANRRR--GFAAFTPHTAARYOGR-----VVIDEAPSLP--PHILLH--MQR	1045
Qy	698	SSRIVLVDVHQIGPIDFOGTSANMPLVBDVVKQCRRTFNOTKRCSPADVATTFFOSLY	757
Dh	1046	AATYHLLADPQOIPAIIDEHAGL--VPAIRPDIGP--TSWVHTYHMPDV--CELIKAY	1100
Qy	758	PGCTTSGCVASISHVADYRNSOAOQLCTPOEBSRHAAGAMTVHEAOGRTPASVILH	817
Dh	1101	PMIQTTSVRLSLFWGEP-----ANGQKLVFQOAKPAN--PGSVTHAEOGATYETITII	1154
Qy	818	YNSGTAEOKLIAEKS--HLVIGTRHYNHLYIRDTGDIEROLNHSAAEYFTDIPALE	875
Dh	1155	---ATADARGLIGSSRAHAIVALTRETEKCVIIDAPG--ILNEBGIS--DAIVNFEIAGGE	1209
Qy	876	ITTVKPSBEVGRN-----EWMATIPROGATFHGAHLKRNFGQPDGCVALKATGEV	930
Dh	1210	IGHORPSYIPGNNDPANDVDTLAAFPSPCOI--SAHQIALEELGNRP--VPAVAIVLPPCEL	1266
Qy	931	FGGRAKIVLELAEPPDATKPHRAFO--EGVQWVKVYNASNKHOQLOTLLSRYSKADPL	989
Dh	1267	EOGLLYLPQELITTCDSVY-----TPELTIIYHNCMAAPSORKAVLSTLVGRIGGR--KL	1319
Qy	990	HEAK--EDVKMLNLSIDRHMWTVEBDARVETOLTQKFTORGCTVEDLLEPD--DPYIR	1046
Dh	1320	YNASHSDVR-----DSLARFIPAIGVQVTTCELYELVEAMVEKGOGSAYLEIDLCNBDVS	1376
Qy	1047	DIDPLMKTQOK--VSKKPIINTGRVGGIAAHSGSLNPLVLAAMTRIIEE--ILTGSATYRYS	1104
Dh	1377	RITFEQOKCNFTTGETTIAHGKVGOGISAWMSTFCALGPMFRAIEKAILALLPGSVFYG	1436
Qy	1105	NGLPDEE--EAMULEKINQVPHATVSAWMTVEPDHANNTSELEFPAALIRIGTPAAAVN	1163
Dh	1437	DAFDPTVSAVAALAAKAMV-----FENDPSEFSDTQNNFSLGLECALIMEBGHOMLIR	1499
Qy	1164	LFRERCGRTLRKAGLGSVEVDGL-----LDGSAAMTBCORNTIFSAAVMLTL--FRGYKF	1216
Dh	1492	LY-----HLIRSAMILOAPKESLAGFWKSGSGEBOTLLMNTVMMAVITHCYDPRDPOV	1545
Qy	1217	AAFKDDSL--LCSHYLAFPDASRLMGERYKTKHLKVBQKVPVITGLIVSAEOVVL--DP	1274

QY 165 FKSRY-GIANSYLDVLEELANAFENHGLHMDPAFMHMBELLMDNVNVAELGIRPHV 223
 Db 193 YTGQVVAIALHSYIDIPADEFGAALLRQNVHVCYAAFFSNNLLLEBDSHVLD----- 245
 QY 224 IEBMAVYKDCAFQGGDLRHPBELDPINESQERRIERLAAGSYSR-----RAVIFSGDD 278
 Db 246 -----EINACFORQDGD-RLTF-----SFASES-----TLNYSHSYSNLLKYCKTYFPASN 290
 QY 279 DWGDAYLHDF-----HTWL-----AYLVR-----NYPTPEGF 306
 Db 291 --REVYKKEFLVTRVNTWPFCKFSRIDPFLYKGVAAHGVDESGFYKAMEDAMHKKTLLAM 348
 QY 307 SLHIEVQRHGGSSIELAITRAPPDGRMLAVYPTRSQGLCPNIFPYADASGTHKTLIT 366
 Db 349 CNSERILLEDDSSSVNWF-----PKORDMVIPLFDISL-----ETSKTRKKEVLV 394
 QY 367 SOHKVNNLNMQRPEKELVDMTVLMSFAPARLARATVVAEVTSSWNISPADLVRTVV 426
 Db 395 LKDFVYVVLNHIKITYOKAKALTYSNVL-SFVESIRSRVINGVTRASENDVDKS-LTQSLS 452
 QY 427 SLVYLIH-----IERRAVAVAKTAKDVFGETS--FWESLKHVYLGSCGGRNLKGT 476
 Db 453 MTFELHKLAVLKODLLISKFPALPKTVSCHVWDEISLAFGNAPPSIKERLIINKKLIT 512
 QY 477 D-----VFTKRVVDKTVVHSLGDIICVRLSPBOGVGLPSRVPARVPHDRE 525
 Db 513 ENALEIRPDLVYTFHDLVSEYIOM-SVDMPEVLDIRKKMEET-----EEMYNALSB 562
 QY 526 LEVLEAGCVN-----ERP----- 539
 Db 563 LSVLKNSDKPEVDVFSQCSQLEVDPMATAKVIIVAVNSNSGLTLTTEQPTANVALALQ 622
 QY 540 -----VPSTPVEEPQ-----GFDADL-----WH-A 559
 Db 623 DSEKASGALVYTRSDVBEPSIKGSMARGELQLAGLSDVSESSYTSSEIESLEQTHMA 682
 QY 560 TAASLPEYR-----ATLQAGLNTDVQOLK-----ITLENALK- 591
 Db 683 TASSLHKOMCSIVYTGPLKQOMKNFIDSLVSLAASVNLVILKDTAALIDLETQKF 742
 QY 592 -----TIGG-----LTISP----- 600
 Db 743 GULDVASKRMLVKPSAKNHANGVETNARKYVALLLEHDEFGIITCDWRERVAASPSV 802
 QY 601 -----VRGLE-----MYEGRPSSGKTGTLIALLEAAGKALVYAPTR 638
 Db 803 YSDMAKRTLRRLKDGEPHVSASAKVVLVDGVPCKGTKEILSRVPE--BDLLIVGCRQ 860
 QY 639 LREAMDR-----IKPPSASATQVHALAILRRATAGAPATVVIDECFMPFLVYVAIV 693
 Db 861 AAEMIRRRANASGIIVATKDNVTRVDSFLMNYGKARQCFRLFDIEGLMHTGCNMFV 920
 QY 694 ALSSPSIVLVGDHOGFID-----FOGTSANMPLVDVYQCRRTFNQKCRPAVVA 749
 Db 921 EMSLSDIAYVGDQOIPYINRVGTGPPYPAHFALEVEDEG--TRRT--TLRCPADV-- 973
 QY 750 TTFQSLVPG-----CTT-----TSGCVASISHVADPYRNSAQTLCTFOEKS---RHGA 797
 Db 974 THFLNQVEGHVMTCSSEKSVSGBEMWSGAASINPVSKPLGKLITFQSKEXALLSRGY 1033
 QY 798 EGAMTVEAQRTPASVILHYNGSTABOKLAESKSHLLVGITRHTNH-----YIRDP-- 850
 Db 1034 TDVHTVEVQGETYADVSLVLTPTPVSIIARDSPHYLVISLRTKSLKYTYVVMDFVLS 1093
 QY 851 -----NGDI-ERGL----- 858
 Db 1094 IIRDLLEVSSYLLDMYKVDAGTQXOLOVDVFKPNPLFVATPKTGDISDMQFYDCKLPG 1153
 QY 859 -----NHSKAEVFTDIPAPLEITTVKPSSEVOBN-----VMAITPQSATP 901
 Db 1154 NSTLLNNYDAVTMLCTDLSLVNKOICIDMSKSVAAAPKDAKTLIPMRTAAEMRQGLL 1213

QY 902 HGAHILNRKPGDPOGCGYALAKTGYEVGGRAKINVE-----LAEPDAPTKPRAP 954
 Db 1214 ENLVAMIKRNFNSPELSGVVDIENT-----ASLVVDKFPSSYLKE--KRKPKNF 1262
 QY 955 QEGVQWKTVMASNKHOALQTLRSYTKRSADLPLEHAKGDVKMLNSLRHMDWTBED 1014
 Db 1263 -----SLFSR-----ESLNR--WIAKOR 1278
 QY 1015 ARDAVETQLKFTQRGTYEDLLBPDPYIRIDPLAKTQ--QKVPKPDINTGKVGQGI 1072
 Db 1279 -----QVTI-----GQLADPFDVLPVAVDQGRHMKIQRKQKDLISQTEVPALQTI 1325
 QY 1073 AASKSLNPLAAMIRLEBILR-----TGSRTVRSNGLPDEBEMMLBAKINQVPA 1126
 Db 1326 VYHSKKINAFGR-----LFSSELTRQLDSDISSRFLPFTRTKTPQOIEPFDLD----- 1375
 QY 1127 TFSADWTEPDTANNRSELPALAE-----RIGTPAAAVNIPREBCGRKTLPAKGLGS 1181
 Db 1376 SHYPMVDLELDVSKYDSQNEFHCAVGEIWRRLGLEDFLAEVWKQGRKTKTLKDYTAG- 1434
 QY 1182 VEVVDGLL-----DSGAANTPCRNITPSA--AVMLTFRGVYKFAFKGDSILC--GSHY 1231
 Db 1435 --IKTCLMYQKSGDVTTPIGNTVYIASCLASMLPWEKLIK-GAFQGDSDSLVPRKGBY 1491
 QY 1232 -----LRPDASRL-----HMGRRYKTKH-----LVBYOKIVPYIGLVSAQOV 1271
 Db 1492 PDIOAANLMMNFBALKFKQYQYFCGRYVYHNDRCIVYVDDPLKISKIGAKRIKOWDH 1551
 QY 1272 LDPVSALIKIFGRCTSELT-----YKXYEAVADITKGSMDAR--YHSLICHS 1319
 Db 1552 LEEFRSL-----CNVASLNCAYITQLDDAAGEVHTKTPSPGTVYSKVLKYL 1601

RESULT 6
 RPO TOML
 ID RPO TOML STANDARD; PRT; 1616 AA.
 AC P03587; O41352;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RNA-directed RNA polymerase (BC 2.7.7.48) (189 kDa protein) [Contains:
 methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
 OS Tomato mosaic virus (betarain I) (TOMV) (TMV betarain tomato).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 NCBI_TaxID=12252;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85157522; Pubmed=6549393;
 RA Okada Y., Aoyagi M., Yamashita Y., Saito H., Ikawa S., Meshi T.,
 "Nucleotide sequence of the tobacco mosaic virus (tomato strain)
 RT genome and comparison with the common strain genome.",
 RL J. Biochem. 96:1915-1923 (1984).
 CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 RNA REPLICATION.
 CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
 CAPING AND AN RNA HELICASE.
 CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 [RNA] (N).
 CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
 BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
 CC -----
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC EMBL; X02144; CA26085.1; --
 DR EMBL; X02144; CA26082.1; --
 DR InterPro; IPR001786; RNA_dep_RNApol12.

RRPO_TOMK2
 ID RRPO_TOMK2 STANDARD: PRT, 1616 AA.
 AC P89676; P90349;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
 Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
 OS Tomato mosaic virus (strain Kazakh K2) (TOMV) (TMV strain K2).
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 NCBI_Taxid=138312;
 RN (1)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=98012318; PubMed=9454068;
 RA Belenovich E.V., Genetozov E.V., Novikov V.K., Zavrlev S.K.;
 RT "Properties and structure of the tobacco mosaic virus strain K2
 genome";
 RL Mol. Biol. (Mosk) 31:826-830(1997).
 CC -!- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 CC RNA REPLICATION.
 CC -!- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
 CC CAPING AND AN RNA HELICASE.
 CC -!- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC {RNA} (N)
 CC -!- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
 CC BETWEEN CODONS FOR GLN-1116 AND GLN-1118.
 CC -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 DR EMBL: Z92909; CAB07439.1;
 DR EMBL: Z92909; CAB07438.1;
 DR InterPro: IPR001788; RNA_dep_RNAPol2.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSVir.
 DR InterPro: IPR002588; V_methyltransfer.
 DR InterPro: IPR000606; Viral_helicase1.
 DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransfer; 1.
 DR Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
 KW CHAIN 1 1616 RNA-DIRECTED RNA POLYMERASE.
 FT CHAIN 1 1116 METHYLTRANSFERASE/RNA HELICASE.
 FT NP_BIND 833 840 ATP (POTENTIAL).
 SQ SEQUENCE 1616 AA; 183614 MW; C2F5CE0C8C965336 CRC64;
 Query Match 3.6%; Score 319.5; DB 1; Length 1616;
 Best Local Similarity 19.5%; Pred. No. 3.9e-11;
 Matches 315; Conservative 133; Mismatches 521; Indels 587; Gaps 72;
 QY 24 QORAVKLDFAAPKALETL--HRLYYPLFKGGLPPTQHP--LAGHQRVAB----- 72
 DB 42 RDRRPKNVSKVSEBQTLIAKAYPEFOI--TFYNTQNAVHSLAGLSLELEYLMMQ 98
 QY 73 -----EVLNFA-----KGRSTVLEIGPSLSALKHGAPAPADYHGGCTKGTTR- 119
 DB 99 IYGSUTYDIDIGNFASHLFGRAVY-----HCCN-----PRLDVRDI--MRHQKDS 144
 QY 120 -----GSRHITALESRSVATGRPEFKADASLLANGIASRTFCVGVGSC-----A 164
 DB 145 IELVLSRLERGNKHV-----PWFQKEAFERYAMENREVVCHTTPQTCRHSQSC 192
 QY 165 FKSRY-GIANSLSYDTLEELANAFENHGLHVRAPWMPPEBLLYNDVNAALGVRFFV 223
 DB 193 YNGRVVAIALHSIYDIPADEFGAALLRKVVHVCYAAFHSENLILEDSHVNDL----- 245
 QY 224 IEPMAVKCAFGQGDRLHPPELDINESQBRIRERLAARGSYR-----RAVIFSGDD 278

DB 246 -----EINACRQDGD-RLTF--SFASRS-----TLNHSYSNLTLYCKTTPRASN 290
 QY 279 DWGDAYLHDP-----HTWL-----AYLVR-----NYPTPGPF 306
 DB 291 --RBYVMKEFLYRNVNTPCKPSRIDTFLFKGVANHGVDSEQFYKAMEDAMHYKTLAM 348
 QY 307 SLHIVQRRHGSIELRTTRAPPGDRLMAVPRRSQGLCRIPNIFYVDAGSTENKTLIT 366
 DB 349 CNERILLEDSSVWYWF-----PKQRDMVIVLPDISL-----ETSKRTKEVLV 394
 QY 367 SOHKVNMILNFMQTEPEBGLVDMYTLMSFAPARLRAIYVASEVTSSWNISPADIVRTV 426
 DB 395 SKDFVYTLNHRITQALTYSNVL-SFVESINSRVTIINGVTARSEMDVDS-LLQSIS 452
 QY 427 SLVYLHI-----IERRAVAVAKTADKDVFGETS--FWESLKGVLGSCCGLEMLKGT 476
 DB 453 MTFPLHTKAVLKODLLSKFALPGKTVSGYVMBEISLAFGNAPFSIERLINRLIKIT 512
 QY 477 D-----VETKRVVDKTRVHSLGDIICVRLSPBOYGLPSRVPPARVPHRE 525
 DB 513 ENALEIRVPDLVTFHDLVSEYKM-SYDMVPLDIRKQMEET-----EEMYNALSE 562
 QY 526 LEVLEBAGCVN-----RRP----- 539
 DB 563 LSVLKNSDKFDVDVFSQMCQSLBVDPMTAAYIVAVNSBGLITTFBQPTBANVALKQ 622
 QY 540 -----VPSPPVEPQ-----GPDADL-----WH-A 559
 DB 623 DEKASDQALVYTSNDVBERPISRGSMARGELQLAGLSGDVBESSYTRBETIESLGFHMA 682
 QY 560 TRASLPERR-----ATQAGLNDVYQK----- 583
 DB 663 TASSLIHQMCQSIYVTGPKLVQCKMNPIDSLVASLSAAVSNVLKTLQTAIAIDETROKF 742
 QY 584 -----ITLEN----- 588
 DB 743 GVLVDASRMLVYKSAKHNANGVETTKARKYVALLBEDRGILLICDMRRVAVSSESV 802
 QY 589 -----ALKTI-----DGLTSPVRGLEMYBGRPGSGKTGLTALAAAGKALVYAPTR 638
 DB 803 YSDMAKLTARLRLKDGSPHVSAAKVAVLDGPGCGKTKELSRNPF--BDLLVPGQ 860
 QY 639 LREANDR-----LYPBAATQVYALAILRRATLBAAPATVVIDCEMPPLVYVAIVH 693
 DB 861 AAEWIRRRANSGIIVAKDKVNRVTDSEFLMVGKARCOFKRLFDDEGLMLTGCVNPLV 920
 QY 694 ALSPESRVLVGDVHVGIFID-----FGTSANMPLVRDVVQCRRRTNQTGRCPADVA 749
 DB 921 EWSLCDIAYVGDVQIYINRYVTFGFPYPAIPAKLVDVEV--TRKT--TLRCFADV-- 973
 QY 750 TTFQSLVPG--CTT-----TSGCVASISHVAPDYNSQAQTLCTQBERKS--RHGA 797
 DB 974 THFLNQRREGHYMCTSSSEKKSQVSGEMVGAASINVSRLKILITFQSDIBALLSNGY 1033
 QY 798 BGAMTVHQAQRTFASVYLATNGSTABOKLLABKSHLLVGTTRTNHL-----YIRP-- 850
 DB 1034 ADVHTVHVEQSETYADVSLVRLTPPVSIIRADSPPHVLVSLSRHTKSKLYTIVVMDPLVS 1093
 QY 851 -----TGDI--BRQL----- 858
 DB 1094 IIRDLERVSYLLDMYKVDAGTQKQLQVDSVFKAFNLFVAAPKTDIDSDMQYYDKCLPG 1153
 QY 859 -----NHSAKAEVTDI-----PAPLEITTVKSEBAYORNEVMAITIPQ 897
 DB 1154 NSTLLNNTDAVTMKULTISLVNKCDCILDMSKSVAAFKD--VKPT-LIPMRTAAEMRQ 1209
 QY 898 SATPHGAILHLKAKPGDPPDGCVALATGYEVRGAKINVB-----LAEPATKXP 950
 DB 1210 TGLLENLVAAMIKNNSPDELGVVDIENY--ASLVVDKFPDSYLLKE--KRPK 1258
 QY 951 HRAFOGQVQWVYVTAANKQALQTLISRYTKRSADLPLHBAKEDVYKMLSLDHEMDT 1010
 DB 1259 KKNP-----SLFSR-----ESLNR--WI 1274

QY 1011 VTEDARDAVETOLKFTORGCTVEDDLEPPDPYTRDIDFLMKTQ--QVNSPKPINTGKV 1068
 Db 1275 AKOE-----QVTTI-----GQLADFPVDLPADVQYRHMKAPKPKLDSLIGTETXPA 1321
 QY 1069 GGGIAAHSKSLNFVLAAMRILEBILR-----TGSRTVRYSGNGLPDEEAMLEAKINO 1122
 Db 1322 LQTIYHSHKKNIAIGP---LFSELTRQLDSIDSSRLPFRKTPPAQIEDPFGDLD--- 1375
 QY 1123 VPHATFVSADWTEFPTAHNNTSELLFALLR-----RIGTPAAMVNLFRERCGKRTLRK 1177
 Db 1376 ----SHVPMVDLDELVSXKYSQNEFHCAVEYIWRRLGLBEDFLAEVWKQGRKRTLLDQY 1431
 QY 1178 GLGSVEVUGL---DGSAAWTPCRNTIFSA---AVMLTLFRGVFAAPKGDLSL 1226
 Db 1432 TAG---IKYCLMYORKSGDVTTFIGNTVIASCCLASMLMEKLIK-GAFCGDLSL 1483

RESULT 9
 RRPO_OR SVC STANDARD; PRT; 1612 AA.
 ID RRPO_OR SVC STANDARD; PRT; 1612 AA.
 AC P89659; 039640; (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [Contains:
 Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
 OS Odontoglossum ringspot virus (Isolate Korean Cy) (ORV-Cy).
 OC Virusess; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
 OX NCBI_TaxID=138661;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96381046; PubMed=8789059;
 RA Ikegami M., Isomura Y., Matsmoto Y., Chatani M., Inouye N.;
 RT "The complete nucleotide sequence of odontoglossum ringspot virus (Cy-
 RL Microbiol Immunol 39:995-1001(1995).
 CC - FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
 CC RNA REPLICATION.
 CC - FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
 CC CAPING AND AN RNA HELICASE.
 CC - CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
 CC [RNA] (N)
 CC - MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
 CC BETWEEN CODONS FOR LEU-1112 AND GLN-1114.
 CC - This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D13941; BA021828.1; -
 DR EMBL: S83257; AAB49498.1; -
 DR InterPro: IPR001788; RNA_dep_RNApol2.
 DR InterPro: IPR007095; RNA_pol_DS_P9.
 DR InterPro: IPR007094; RNA_pol_PSVr.
 DR InterPro: IPR002588; V_methyltransfer.
 DR InterPro: IPR006066; Viral_helicase1.
 DR Pfam: PF00978; RNA_dep_RNApol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransfer; 1.
 KW Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
 FT CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE
 FT NP_BIND 826 833 ATP (POTENTIAL).
 FT NP_BIND 1612 AA; 183252 MW; 3F803A2B9611E0DF CRC64;
 SQ SEQUENCE

QY 108 DY-HGC-TKYTRDGSRTITLBSRV-----ATGRPEPKADASILLANGIASRTFCV 157
 Db 121 DYVHCCHPMNLDIRVAVHINOQDVTSTYLARLSRKGLPFVQOSAPFKYNNPDVAVCCD 180
 QY 158 DGVSQCAFK-----SRGIANHSLYDVLLEELANAFENHGLAMVRAFMHMEPEELINDNV 212
 Db 181 KRPODCSYVDLPCKTAVVALHSTYDIPADFGAALRKQDVHICYAAPHSENLIL--ET 238
 QY 213 VNA---ELGYRPHVIBSPMAVKDCAPQGDRLRHPBELDFINSGERRIERLARGSYSR 269
 Db 239 TSAPLDEIGATF-----YKSGD-RLSF--FPQNESTLN----- 268
 QY 270 RAVIFSGDDMGADYLDFTHTLALYLVRNYPFPFSGSLHIEVORHSGSIELRIITRAP 329
 Db 269 -----YESYKNVITYCKTTPPASNRVYHKEFCTRNTWPKCKTKVD 314
 QY 330 GDRMLAVPRTSGGLCHIPNIFFYADASGTEHK-----TILSQHVM----- 373
 Db 315 YFLRQVYTRGEDS-----BOFTYAMDAMWYKTKTLMLKCEKRTIFDRRAVNWPPKVK 369
 QY 374 -----LNFQTRPEKELVDMTVLMSPARRLRA 402
 Db 370 DMVIVPLPDGSVTSGMKKBSVMVWVKDPVTVLNHITTYQDKALTYKNVL-SFVESIRSR 428
 QY 403 IVVASVYTESWNIISPADL-VRTVSLYLHIERRBAVAVKTAQDVFGRSTFWBSLK 461
 Db 429 VIINGVARSSEMDVXSVLQALSWTFLLQTLAELAKQVVLKFKQKEDVTYNLFMKQIS 488
 QY 462 HTVLSGGLRLKGT-----DVPTKRVDKRYRNSLDDIICDV 500
 Db 489 DAVDDL--PSEIKETLISGCFVKAESLQIKTPDEYITFADKLMERYK----- 535
 QY 501 RLSPQGVPLPSRPVPAV---FHDRELELYRAGCY----- 535
 Db 536 --ATEBLQHDIDISPLRAKRYNALSELSTLXKCDSEFDITQFNLCBKHQIDPDVAVK 593
 QY 536 -----NRPVPSPTPYEE-----PQFPDAD-----LWHTAASLPYRATLQAGL- 575
 Db 594 IVALMKBELTLFPKNPPEALSDALSPLKQLDWRPDLKLKSTCAPPSVK-TLDSGLLP 652
 QY 576 -----NTDVKQLK----- 583
 Db 653 KQSYGDERQFESQSVSVDFHLKSVSVKKSNSAVVTGRLVKQMKNYMDLISASIS 712
 QY 584 ITLENALKTI----- 593
 Db 713 ATVENLCKVLKDVGVADPESAKSGVDYVKGKRLLPKKGCHAMGVALNNGERVVLL 772
 QY 594 -----DELTLSPVRLR-----MYSGPPSGK 615
 Db 773 EMADGPPICGDMRRAVAVSSDLISYDMGKLTLLCLCKDGFPVSDAKVTLVDVPGGK 832
 QY 616 TGTILALBAAGGALVAPTRRLREAMDRIKPPSASATQHVALLIRRTABGA----- 671
 Db 833 TKELLETVPN--DEDLILVPGKEACKMIKR-----ANKSGHV-----RATKQNVRTVD 879
 QY 672 -----PEATVVIDECFMPPLVYVALIYHALSPSSRIYLVGVHOGIFDFOGTSANM 722
 Db 880 SPLHLKPKTYNKLPIBEGMLMHTGCMPLIALSHCBAAVVGUTBOIPLFN--RYANF 936
 QY 723 P-----LVNDVVKQCRRTFMQTKCPADVVAATPFPQSLYPS-CTTSGCVAISIHVA 774
 Db 937 PYPGHFATLYVD-----HREVRSLRCPADV--THFNNSKYDKGKVLCTNDVINSVDAEV 989
 QY 775 -----PDVRNSQAQTLCTQSEKSRHAGG-----ANTYHRAQGRTPASVYLH 817
 Db 990 VRGQGVNPSKPLKGIITFTQSDKABLERGEVSTGEBINTVHIEQETEDVSVV 1049
 QY 818 YNGSTAOKLAEKSHLLVGTTRTN-----HLVYRDPGTGDIROLANSA-----KAVFT 868
 Db 1050 RLTFPTPELISKSPPHVLVLTTRTKSPKYSVVLDPVLVKVCSLSKVSBDIILMYKYDA 1109

Query Match 3.4%; Score 301.5; DB 1; Length 1612;
 Best Local Similarity 18.8%; Pred. No. 4.9e-10;
 Matches 307; Conservative 200; Mismatches 570; Indels 553; Gaps 68;

```

QY 869 DIPAPLEITVYKPSSEV-----QRNEVMAITPQSAT---PHGAIHL-LRKNFG 913
DB 1110 GILXQIQVGSIFKGENLFWPCPKSGYISDMQFYDILLRGNSTILNEYDAVTMNLKENNL 1169
QY 914 DOPDCGCVLAKTGVFVPGRAKINVELAPDAPKPHAFQGVQVWVKTANSNGOAL 973
DB 1170 NKVDC-----TIDFSKSVSVPRQOQDEFPIPV---ITTAERPRSRUL 1208
QY 974 QTLISRYTKRSADLP---LHEAKEDVKMLNSLDHMD-----WTVEDARD 1017
DB 1209 LENVLVMIRKRNFSPLTGILDEPTAEIYVN---KFMDAYIIDLSGNVTWMTSDAHH 1265
QY 1018 RAVEFQQLKFTQRGVIEDLLEDDYIDIDPLMKTQ---QKSPKPIVTKGKGQIAAH 1075
DB 1266 RNMWAK-QEKSTI--GQLADFDVFDLPAIQYKIMIAQPKQKDLSPQDEYALQTIYVH 1322
QY 1076 SKSLNVLVLAAMI-----RIIEELIRGTSRTVRVSNGLPDEEAMLEAKINQVPHATFVS 1130
DB 1323 SKQINAIKQPLSELTRQULERI--DSKFLFTYTRKTPQOIEEFF--SDLDSTVPMEVLE 1378
QY 1131 ADWTEPDTANNTSELLFPAALLERIGTPAAAVNLFRERGKRTLRAGLGSVEVDGL--- 1187
DB 1379 LDISKYKQNEPHCAVEYLWEKLG---LNGFLBEVWKQGHRTSLKDYTA-GIKTC 1432
QY 1188 ----LDGSAWTFRCRTTISAAN---MLTIFRGVKAARFGDSSL-----CG 1228
DB 1433 LMYQRKSGVTFITGTVIIMACIASMIPMDKVIK-AAFCGDDSMIYIPKGLDLPDIOG 1491
QY 1229 SHYL-RFDAS--RLHNG---ERYKTHLKVQKIVYIGLAVSABQVLDVPSALKIF 1282
DB 1492 ANLMWTFEAKLYKRGYFCGRIYIHHDKGAT-----VYIDPKLISK-- 1534
QY 1283 GRCTSELLYSKYVEAVRDTKGSMDARYSHLLCHNSA-----CYNVAPESAAYITDAV 1337
DB 1535 -----LGCRIKIKSLDHLER-----FRISLGDVASLNCAYGQINDAIAVHKTA 1580
QY 1338 VRRGRDPPF 1347
DB 1581 V--NGSFAP 1587

RESULT 10
RPO_ORSVS STANDARD; PRT; 1612 AA.
ID RPO_ORSVS STANDARD; PRT; 1612 AA.
AC Q84133;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-directed RNA polymerase (EC 2.7.7.48) (183 kDa protein) [contains:
DE Methylyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
OS Odontoglossum ringpot virus (isolate Singapore 1) (ORSV).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirinae.
ON NCBI_TaxId=138662;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96257213; PubMed=8666266;
RA Chung C.G., Wong S.M., Mahant P.H., Loh C.S., Goh C.J., Kao M.C.C.,
RA Chung M.C.M., Watanabe Y.;
RT "The complete sequence of a Singapore isolate of odontoglossum
RT ringpot virus and comparison with other tobamoviruses.";
RL Gene 171,155-161(1996).
CC -| FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -| FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPING AND AN RNA HELICASE.
CC -| CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC [RNA](N).
CC -| MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR LEU-1112 AND GLN-1114.
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its

```

```

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL, U34586; AAC55012.1; -.
DR InterPro; IPR001788; RNA_dep_RNApol2.
DR InterPro; IPR007095; RNA_pol_DS_Ps.
DR InterPro; IPR007094; RNA_pol_PsVir.
DR InterPro; IPR002588; V_methyltransferase.
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF01443; RNA_dep_RNApol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01660; Vmethyltransferase; 1.
KW transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
FT CHAIN 1 1112 METHYLTRANSFERASE/RNA HELICASE.
FT NP BIND 826 833 ATP (POTENTIAL).
SQ SEQUENCE 1612 AA; 183194 MW; 38444670B0ACB189 CRC64;

Query Match 3.4%; Score 300; DB 1; Length 1612;
Best Local Similarity 19.3%; Pred. No. 66-10;
Matches 285; Conservative 176; Mismatches 535; Indels 484; Gaps 61;

QY 108 DY-HGC-ITKGRDSSRHTALERSV-----ATGRPKADASLLANGISRTTCV 157
DB 121 DYVHCMPMLDTRDVAHHINODTVSTYALARLERSKGLPVQGSAPFNKYMNDPDAVCCD 180
QY 158 DGVGSCAFKSRV-----GIANSLYDVTLEBLANFENHGLMVAFMMPELLYMDNV 212
DB 181 KRFQDSYSGLPKGTIVAGLHSIYDIPADEGALLKRDVICTAAHPSNLL--ET 238
QY 213 VNA--ELGKRPVIEBPMAVNDCAFQGGDLNHPPELDPNFSQERRIERLARAGYSR 269
DB 239 TSAPLDEIGATP-----YKSGD-RLSP--FQVRSSTLN----- 268
QY 270 RAVTSGDDMDKDALHDHTWLAIVLRNVTTPPGFSIHIVQRRHGSIELRTTRAP 329
DB 269 -----YHSYKNVYIKVCKTFPASNRFVYHKEPMCTRVMTWPCKEFKYDT 314
QY 330 GDRMLAVVPRSTQGLRIPIFYVADASTGK-----TILTSQHKNM----- 373
DB 315 YFLFRGVYTRGDS-----EPTTAMDAMEYKTKTLAMNSERTIFRDAAVNFMFPKVK 369
QY 374 -----LTMQTRPEKELVDMTVLMSPARARLRA 402
DB 370 DMVIVPLFDGVTSGMKRSVMVNNKDPFYTVLNIHRTYQDALYQVNL-SFVESISR 428
QY 403 IVVABSVTSSWNISPADL-VRTVSVLYLHIERRAVAVKTAQDVFGSTFWESLK 461
DB 429 VINGVTARSEMDVVKSVLQALSMTFLLQTKAKAKDQVVLKPKFQFDVTNLFWKQIS 488
QY 462 HTVGS---CCGLRNKGTGVVFTK-----RVYDKRVHSLGDIICDVRSLSPVOGFLPSR 513
DB 489 DAVGLFPIKRIKRLISGGFVKVABGSLQIKTDEY-ITPADLWMEYQ-ATEISLQHLDS 546
QY 514 VPPARV--FHIRELEVLRAQCY-----NERPVVS 542
DB 547 KPLERARKYVNLSELISVLKCDPDIQFKMLCEBKIDAPDVAKVIPIPKNBLTLFP 606
QY 543 TTPVER-----PGQPAD-----LMAHTASLRYRTLQAGL----- 575
DB 607 KNTPPALSDALSPKDLMDPCLLKSTCAPPSVK-TLDSGLPKQSYGDERQFSQ 665
QY 576 -----NTVQKOK-----ITLENALKTI----- 593
DB 666 SVVSVDFTLKSVESVTKMSGSAVYTGKTKQKKNVMDYLSAISATVSLCKYLVKOV 725
QY 594 -----DELTL----- 598
DB 726 YGADPESAEKSGVDVVKGMKLLKPKDKCHANGVABLNNGRKVIYLLLEWADFPICGWR 785
QY 599 -----SPVRG-----LEWYBGRPGSGKCTTLAALBAAG 628

```

```

Db      766 RYAVSDSPSYDMGKLTQLSLCKDGEVPLAKPKTLVDYFGCKTKEIETVNF--D 843
      629 KALYVAPTRELBRANDRIKPSASATQVALAILRAFAEGA-----PFAI 675
      844 EDLIVPGKACKMIKIR-----ANKSHV-----RATRDVNTVDSFLMHLKPKTYNK 892
      676 VVIDCEFMFLYVAIVAHLSPSRSRVLVGDVHQIGFIPQSTASAMPLVDVVKOC--R 733
      893 LFDIGSLMHTGCNVFVALSHCREAMVGDABEQIPFIN--RVANFPKPKFRYTCLVH 949
      734 RRTFNQTKRCPADVVAATTFQSLYLPQ--CTTSGCVASISHVA-----PYRNSQAOT 784
      950 REVRRLSLRCPADV--THPMNSKYDQKULCTNDVIRSVAAEVYRGKGVNPKSKPLKGI 1007
      785 LCFTEEKSRRHGAEG-----ANTVHAQGRTPASVILYHNGSTAEQKLAESKSHLL 835
      1008 IFTQSDKALKERGEVEVSTFGEIINTVHEIQEFTEFEDVSVVRLTPTPLSLSKSPHYL 1067
      836 VQITHTN---HLYRDPTGDIEROLNHA-----KAVPFDIPAPLEITTVKSEEV- 885
      1068 VALTHTHTSKYYSVVLDPVAVKWSDSLSTVSDPILDMYVADGILKQLQVGSIFKGENLP 1127
      886 -----QRENVWATIPQSAT---PHGAHL-IRKNFGDPQPCGCVALAKTGYEVF 931
      1128 VPCPSGYISDMQTYDTLVPGNSTILNEYDVAVTMLRENNANVDC----- 1174
      932 GGRANINVELAPDAPTPKPHRAPOEGVQVKTVAASNKQALQTLISRYTKSADLPHE 991
      1175 -----TDFSKSVSVPRQOEFFT-----PAHRTAER--PRSAGL-LEN 1211
      992 AKEDYKRMNSID-----RHWD-----WTVEADRADRAVEFT 1023
      1212 LVAMTKRNNSBDLGLIDIEDTAEI.VNKKFMDAYIIDEISGCVTPMTSDAFHRRMAC- 1270
      1024 QLKFTORGSTVEDLLEPDDPYRIDIDFLAKTQ--QKVSFKPINTGKVGQGIASHSKSLNF 1081
      1271 QEKSTI--GQLADFDVVDLPAIDQYKMKAKQPKQKGLSPQDEYALQTTIYHSHQINA 1328
      1082 VLAAMI-----RILEEILRTGSTRVYVNGLPDEEBAMLEAKINGVPHATVSAWTFE 1136
      1329 IKGPLFSELTRQLEKTI--DSKFLFYTKTPEQIAFF--SOLDSTVMEVLEIDISKY 1384
      1137 DTAHNTSELPAALIERIGTPAAAVNLPRRCGKRTLPAGKIGSYEVNGL-----LD 1189
      1385 DKSQNEFHCAVEYLIMEXLG-----LNGFLBEVWKQGHKRTSLKDYTA-GIKTCIMYQKR 1438
      1190 SGAAATPCGNTIFSAV---MLTLPRGVKFAAFKGDSSL 1226
      1439 SGDVTTFIGNTVIIMACIASMIPMDKVIK-AAFCGDDSL 1477

```

```

CC      RNA REPLICATION.
CC      -I- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC      CAPPING AND AN RNA HELICASE.
CC      -I- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC      {RNA} (N).
CC      -I- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC      BETWEEN CODONS FOR GLN-1103 AND GLN-1105.
CC      CC
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
CC      CC
CC      EMBL, U30944; AAB0599.1; -.
CC      PIR, S65053; S65053.
CC      DR      InterPro; IPR001788; RNA_dep_RNApol2.
CC      DR      InterPro; IPR007095; RNA_pol_DS_P8.
CC      DR      InterPro; IPR007094; RNA_pol_PSVir.
CC      DR      InterPro; IPR002588; V_methyltransf.
CC      DR      InterPro; IPR000606; Viral_helicase1.
CC      DR      Pfam; PF00978; RNA_dep_RNApol2; 1.
CC      DR      Pfam; PF01443; Viral_helicase1; 1.
CC      DR      Pfam; PF01660; Vmethyltransf; 1.
CC      TRNAsease; RNA-directed RNA polymerase; Helicase; ATP-binding.
CC      KW      CHAIN 1 1597
CC      FT      CHAIN 1 1103
CC      FT      NP_BIND 823 830
CC      FT      VARIANT 919 919
CC      FT      VARIANT 1286 1286
CC      FT      VARIAT 1286 1286
CC      SQ      SEQUENCE 1597 AA; 181621 MW; 7256A908BD3308F2 CRC64;

```

Query Match 3.4%; Score 298.5; DB 1; Length 1597;
 Best local similarity 19.0%; Pred. No. 7.3e-10;
 Matches 287; Conservative 184; Mismatches 524; Indels 515; Gaps 60;

```

QY      171 IANSLYDVTLEELANAFENHGLNVRAPFMNPEE-LLYMDNVNAELGYRPHVIEBPMA 229
      199 IALHSIYDIPADEFGAALLRKVKIKCYAAPHENMLDCCSDVLEIDIGATQ----- 251
      230 VMDCAFGQDGLALHPELDFINESQBRRIERLAAGSISRRRAVTSGGDDMDGDAYLHPH 289
      252 -----RAGD-KLNF--PFHNESTLN-----YTHSES 274
      290 TWIAYLVVANYPTPGFSLHIEVQRHGSIELRITRAPPGDRMLAVVPRTSQGLCRIPN 349
      275 NIIVYCKTFPPASQRYVYHGF-----LTVRVNMYCKEPR 311
      350 I-----PYTA----- 354
      312 VDTFLFGVYKTSVDSSEFYKAMDAMEYKKTILANLNSERTIPROSAINFMFPKVRDM 371
      355 -----DASGT-----EHTITLSQKRVNMLNFMQTRPEKELVDMTVLMSPARALLRAV 404
      372 VIIPLFDASITTGRRSRRLVNVKDFYTVLNIHITYQAKALTYANVL-SFVESLRISRYI 430
      405 VASRYTESGWNISPADLVATVTVSLVLIHIERRAAVAVK-----TAKDVPFGSTSPW 457
      431 INGVARBSMDVDAKILQPLAMTFFLVKLSHVQDEIVLKKQKQKDATKELIM--SLIC 488
      458 ESLKVLVAGSC-----CGLRNKGTDV-----VPTKRVVDKXR----- 489
      489 DALKGIVISVKETLARGFVKLAESLEIKIPELYCTFEDRLVLYKRTBEPQSCDLSKP 548
      490 -----VNSIGDIICDV-----RLSPBQVG-----FLPSRV 514
      549 LBSSEKTYNALSELVLENLGSFDLDAPFKELCQKKNVDPDAKAVVVALMNSLTLPPFK 608
      515 PPARVPHDEELEVLRAG--CYNERPVSPTRPEE-----PQG-----FDADL 556
      609 PTBEEVAALSGEVVQDEGLRLSNKAPPCVSNLKGVLVPAQGLCPNGANPRVMDISE 668

```

```

QY 557 WHATAA-----SLPEYRATLQAGLNTDVK-----580
D 669 FHLKSVDAVKKGAMMSAVYTGKIKVQGMKNVVDLSLSATVENLCKVLRDVHGVDPS 728
QY 581 -----OLKTTLEN-----588
D 729 QKSGVMDVRRGWLKPNKACHAMGVAEDANHLCLVVLNMBGKPVCEBETWRLAVSS 788
QY 589 -----ALKTI-----DGLTSPVGRLEMEYGPSSGCTGTLIALAASAGKALYVA 634
D 789 DSLVSDMGKLTITSCCRDSEPEPFAKVLVDGVGCGTKTEIKVPS--EDLVLV 846
QY 635 PTRELEAMDRRIKPPASATQVALAILRRATAEGA-----PFATVVIDEC 681
D 847 PKKASGMIIR-----ANQ-----AGITRADKONRTVDSFLMHPKPVFKKLPIDEG 895
QY 682 FMFPLVYVALVHALSPSSRIYVDVHQIGFIDQGISANP-----LYRDVVKCCR 734
D 896 LMLHTGCVPNLMLSHCDVAVVYVDTOOIPFI---CRVANFPYPAHFAKLVVD--EKEDRR 951
QY 735 RTFNQTRCPADVATTFPOSILYGP-----CTT-----TSGCVASISHVAPYRSQAOTL 785
D 952 VTL-----RCPADV---TFPLNOKYDGSVLCSSVERSVSAEYVRGKALNPITLPLEGKIL 1005
QY 786 CFTQEEKSR--HGAEQAMTVHEAQRTFASVILHYNGSTAEOKLAEKSHLVIGTRHT 842
D 1006 TPTQADFFELLDKQYKQVNTVHEVQGETEYKTAIVRLATPFLLEISASPHVVALTRHT 1065
QY 843 N---HLIYNDP-----TGDIRQJNHSKAAVEFTDIPAPLE-----ITTYKPS 882
D 1066 TRCKYTVTVLDPVNVNVLSELGKSNFLMEYKVESGTOXQOQIDITVFKGTLNLPPTKSG 1125
QY 883 EEVQRNEVMATIPQSAFPHGAILLRKNPFD-----QPDCCVALATGYEVFGRAKIN 938
D 1126 DMRDMQFYVDTLRLGNSITILNEPFAVTNMLRDLISLNVKDC-----R 1166
QY 939 VELAEPTATPKPHAFQEGVQWVVTNANSNGHQAQTL--LSRYTKESADLPLEHAEDV 996
D 1167 IDFKSKVQVPERVFNK-----PLRTAEMPRTRAGLEENLVAMIKNNMNPDLGTITDI 1222
QY 997 KRLMN--SLDRHMDWTVED-----ARDRAVFETQKFTQRGGTVEDLE--PDDPYI 1045
D 1223 EDTASLVEKFWDAVYVVEFGTGMAMTRSFGRMLS--KQESSVQGLADFNFDLPAV 1281
QY 1046 RDIDFLMKTQ--QKVSPPKINTGKVGCGIAHSHKSLNVLAAWI-----RIIEELIRGSR 1099
D 1282 DEYKMTIKSQPKQKLDUSTODEYALQTTVYHSHKKNALIFGPMFSELTMLLETITD--SK 1340
QY 1100 TVRYSNGLPDEEBAMLEAKINQVPHATFVSADWTEPDTAHHNTSELFFALLERIGTPA 1159
D 1341 FLVFTTRKTPTQIEEPFSDLSQA--MEILELIDISKYKSKQNEPHCAVEYKIMKLGIDD 1398
QY 1160 AAVNLFRERCCKRTLLRAKGLGSVEVDGL-----DSGAAMPCCRNITFSAV--MLTFR 1212
D 1399 WLAEVWRQGHKRTLLKDYTAG--IKTCLMYQKSGDVTFIQTIIIIAACLSMIIPMDK 1455
QY 1213 GVKFAAPRGDLSL-----CGSHYLAFPDASRL 1239
D 1456 VIK--AAFGDLSLYIPKGLDLPDIQAGANLTMNFEALFKKCYGFPGGRVVIHHRGAI 1514
QY 1240 HMGGRYKTKHLKVEQKIVPYIGLLVSAQVLLDVRSLAKIFGRG--YTSER--LYSK 1294
D 1515 -----VYIDPLKISLKGCHIRDEVHLEIRSL-----CDVTSILNNCAVYSQ 1559
QY 1295 YVEAVRDITK 1304
D 1560 LDEAVAVHK 1569

```

RESULT 12
RRPO_PMMVJ
ID_RRPO_PMMVJ STANDARD; PRT; 1612 AA.

```

AC P99657, P90347,
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-directed RNA polymerase (BC 2.7.7.48) (183 kDa protein) [contains:
DE methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
OS Pepper mild mottle virus (strain Japan) (PMMV-J).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI TaxID=138663;
RN [1]
RP SEQUENCE FROM N.A.
RA Kirita M., Akutsu K., Watanabe Y., Tsuda S.;
RT Nucleotide sequence of the Japanese isolate of pepper [Capsicum
RT annum] mild mottle tobamovirus (TMV-P) RNA.
RL Ann. Phytopathol. Soc. Jpn. 63:373-376(1997).
CC -1- FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC -1- FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC (RNA) (N).
CC -1- MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UAG OCCURS
CC BETWEEN CODONS FOR GUN-1117 AND GUN-1119.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1eb-rib.ch/announce/
CC or send an email to license@1eb-rib.ch).
CC -----
DR EMBL; AB000709; BAA19167.1; -.
DR EMBL; AB000709; BAA19166.1; -.
DR InterPro; IPR001788; RNA_dep_RNApol12.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR007094; RNA_pol_PSV1r.
DR InterPro; IPR002588; V_methyltransfer.
DR InterPro; IPR000606; Viral_helicase1.
DR Pfam; PF00978; RNA_dep_RNApol12; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01660; Vmethyltransfer; 1.
DR Transferrase; RNA-directed RNA polymerase; Helicase; ATP-binding.
KW CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1117 METHYLTRANSFERASE/RNA HELICASE.
FT NP BIND 835 842 ATP (POTENTIAL).
SQ SEQUENCE 1612 AA; 183303 MW; E7B6C8B8D81P84C CRC64;
Query Match 3.44; Score 297.5; DB 1; Length 1612;
Best Local Similarity 19.64; Pred. No. 8.5e-10;
Matches 318; Conservative 188; Mismatches 542; Indels 575; Gaps 77;
QY 10 RVYAAADVAYANVLCQAVVLDPAFPAKALETL--HLYYPLAPFKGTLPTQHPRI--LA 65
D 30 FLYESA--VEQCNA--HDDRPPVPLRSISEBQTLATAYEFPQI---TFYNTQVAVSLA 84
QY 66 GHQEVAE--EYLHNFARGRSTVLEIGPELSALKILGAPNAPVADY--HGC--TKYGRDGRSR 122
D 85 GGLRSLELEVLMMQIPIPGSTTYIDGKNP--AAHNFPG-----RDYVCCPMNDLRVNR 137
QY 123 HITLSESRVATG-----RPRKADASLLANGIASRTF--CYDGVGSCAFKS 167
D 138 HNAQKDSIBLYSLKQAKKQVIPPYQPCFDKTYDDPOSVVCSPFQHC--BGVSHCTDKV 196
QY 168 RVGIANHSLYDVTLEELANPENNGLAMVAAPMPEMBELLVMDNVNNA--ELGIRFHYIEE 226
D 197 -YVALHSLVDIPDERGALLRRNVAVCYAAPFSSNLLLSYSVSLDDIGAFS----251
QY 227 PMAVDCAFQCGDRLRLFPPELDFINESQERIERLARAGSYSR-----RAVIFSGDDWG 281
D 252 -----RBDGM-LNF-----SFVABS-----TLANTHSYSLVNLKVCTYPPASS--R 290
QY 282 DAYLHDF-----HTWL-----AYLLIVR-----NYPPFGFSLH 309

```

```

Db      291 EYVMEFLVTRNNTFCRPSLDTFVLRGVYHGVGDQFYSAMEDAHYKKTILAMNS 350
      310 IEVORRHSSIELRTTRAPPGDRMLAVPRTSQGLCRIPNIFYADASTEHKILLTSQH 369
      351 ERILLEDESSVVMVF-----PKKDMVIVP-----LEFVSLQNEGRKLAKEMVWSKD 398
      370 KVMMLNMQRRPEPELVDMYVLSFAPARALAIYVASVETSSNMNISPADLVRTVLSLY 429
      399 FVYVYLNIRIYQSALTYVANVL-SFVESIRSRVINGVTAESMDVDKKA-----LLQSLS 453
      430 VLIHIERRAAAYATAKODVFE-----TSF-WESEKGVAGSC-----468
      454 MTFPLQTLAML-----KDDLVAVQKFQVHSHKSLTEYVMDETIAAFHNCFPITKEKLINK 508
      469 -----GLRNKGTD--VFTKRVVDKYRVHSLGDIICDVRLSPEQVGLPSPRPAPV 520
      509 LITVEKALKEIKVPLVYTFHDLRVKEYK-SSVEMPVLDVKKSLDEAEVMYVALSEISIL 567
      521 HDRELE-----VLRBAG--CYNERPVDS-----542
      568 KDSKDFDVDSRMCNTLGVDPVLAQKVMVAVVSNESGLTLTFRPTEANVALALQPTIT 627
      543 -----TPVEEPQGD-----553
      628 SKESGLKIVSDVSESSIKVVRKSEISMGLTGNVSDDEFORSTELSLQCFHNVSTE 687
      554 ---ADLMHATAASLP-----EYRATLQAGLNTDVKKL-----ITLEN-----588
      668 TIRKQNMAMVYTGVLKVGQCKNYIDSLVASLSAIVSNLKIHKDTAIDLETKEKFGVY 747
      569 -----ALKTIDG-----LT 597
      748 DVCLKMLVKPLSKGHANGVWVSDYKCFVALLTYDGENIVCGETRRVAVSSSELYSD 807
      598 LSPVAGLE-----MYEGPQSGKTGLTILALAEAGKALYVAPTRBLRE 641
      808 MCKTAAISVULKQGPPIHSAKVTLDVGPQCGKKEILSRVNF--DEDLVLPQKQAAE 865
      642 AMDRRIKPPSATQHVALLLRATATAGCAPAT-----VVI 678
      866 -----MIRRRANSSGLIATENRTVDSFLMNYGRGCOYKRLFL 906
      679 DECFMFLVYVAIVHALSPSSRIVLVGDVHVGIFIDPOGTSANMP-----LVSDVVKQ 731
      907 DEGLMLHFGCVFLVGMSLCSEAFYGDQIQPYIN--RVATFPYKHLSQLLEVDAVE- 962
      732 CRRRTFNQTKGCPAVVATTFFQSLYPG-CTTSGCVASISF-----VAPDVRNQA 782
      963 -TRRT--TLRCPADI--TFPLNQRYEGQVMCTSSVTRSVSHEVIGQAAMVPVSKPLKG 1016
      783 QTLCTFOEKS---RHGAGAMTVEHAQGRTPASVYLHNNGSTAEOKLAEKSHLLVGT 839
      1017 KVITFTQDCKSLLSRGVEDVHTVEVGETFEDVSIVRLTTPYGIISKQSPHLLVLS 1076
      840 RHTN-----HVIYBDP 850
      1077 RHTSIRKYTVVLAIVSVLRDLCEVSSYILDMYKVDVSTQYQIIESVYKGVNLFVAAP 1136
      851 -TGDI-----ERQIN-HSAKAEVFTDIPAPLEI--TT 878
      1137 KTGVDSDMQYYDKLPGNSITLNEYDAVTAQIIRNSILNVKDCVLDMSKSVLPRESSET 1196
      879 VKPSEEVORNEVMATIPQSATP--HGAILLRKNGFQDQPCGCVALAKTGYEFGC-- 933
      1197 LKP-----VIRAAEKPRKGLLENVMIKKNFNSPELVGVVDIEDTSLVVDKFF 1248
      934 RAKINVELAEPDAPTKPHRAFOEGVQVQVKNVNASKHQALQTLRSRYTKRSADLPLEAK 993
      1249 DAYLIKKEKKPKNIDILSRASLE--RWIEKQKSTIGOL-----ADFPDID-- 1292
      994 EDVKMLNSLDHNMWMTYTEDARDRAVFTOLKFTQRGSTVEDLLEPDDPYIRDIDPLMK 1053

```

```

Db      1293 -----LPAVQY-----RHMKQPK--QR-----LDLSIQ 1316
      1054 TQOKVSPRPINTGKVGQIAHSSKLNFLVA-AMIRILIELRT--GSRTVRYNSGLPDE 1110
      1317 TEYPAL-----QTVVHSHKKNLFGPVSELTROLETIDSRFPFYTKEPTQ 1366
      1111 EAMMLEAKINQVPHATVSDMTPEPDHANNTSBLFPAALLERIGTPAAVNLPRKCG 1170
      1367 IEEFSDSDSN--VP-MDLIELDISKIDSKQNEFHCAVETRIWKRLGDDPLAEVWKQHR 1424
      1171 KRTLRKAGLAGSEVNDGL-----DSGAATPGRNTIFSAV--MLTLFRGVFAAFKGD 1223
      1425 KTLIKDLYAG--IKTCLMYRKSGDVTTFIGNITIIIAACLSMLPWERLIK-GAFGSD 1480
      1224 SLI 1226
      1481 SIL 1483
      1481 SIL 1483

RESULT 13
RRPO PMVVS
ID RRPO PMVVS STANDARD; PRT; 1612 AA.
AC P29058; 084924;
DT 01-DEC-1992 (Rel. 24, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE RNA-directed RNA polymerase (BC 2.7.7.48) (183 kDa protein) [Contains:
DE Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].
OC Pepper mild mottle virus (strain Spain) (PMVVS-8).
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tobamovirus.
OX NCBI_TaxID=31745;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=92113528; PubMed=1765765;
RA Alonso B., Garcia-Luque I., de la Cruz A., Wicke B.,
RA Avila-Rincon M.J., Serra M.T., Castresana C., Diaz-Ruiz J.R.;
RT "Nucleotide sequence of the genomic RNA of pepper mild mottle virus,
RT a resistance-breaking tobamovirus in pepper.";
RL J. Gen. Virol. 72:2875-2884 (1991).
CC - FUNCTION: THE LONGER PROTEIN IS A RNA POLYMERASE ACTIVE IN VIRAL
CC RNA REPLICATION.
CC - FUNCTION: THE SMALLER PROTEIN IS A METHYLTRANSFERASE ACTIVE IN RNA
CC CAPPING AND AN RNA HELICASE.
CC - CATALYTIC ACTIVITY: N nucleoside triphosphate = N diphosphate +
CC {RNA} (N).
CC - MISCELLANEOUS: READTHROUGH OF THE TERMINATOR CODON UNG OCCURS
CC BETWEEN CODONS FOR GLN-1117 AND GLN-1119.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL; M81413; AB02334.1; -
CC EMBL; M81413; AB02335.1; -
CC PIR; J01312; MMTMPV.
CC InterPro; IPR001788; RNA_dep_RNApol2.
CC InterPro; IPR007095; RNA_pol_DS_PS.
CC InterPro; IPR007094; RNA_pol_PSVir.
CC InterPro; IPR002588; V_methyltransf.
CC InterPro; IPR000606; Viral_helicase1.
CC Pfam; PF00978; RNA_dep_RNApol2; 1.
CC Pfam; PF01443; Viral_helicase1; 1.
CC Pfam; PF01660; Vmethyltransf; 1.
CC Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
KM CHAIN 1 1612 RNA-DIRECTED RNA POLYMERASE.
FT CHAIN 1 1117 METHYLTRANSFERASE/RNA HELICASE.
FT NP BIND 835 842 ATP (POTENTIAL).
FT SEQUENCE 1612 AA; 183265 MW; 46567784924802BF CRC64;

```


CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.1sb-sdb.ch/announce/>
 CC or send an email to license@sdb-sdb.ch).

CC -----
 CC EMBL; D63809; BAA09876.1; -
 CC EMBL; D63809; BAA09877.1; -
 CC InterPro; IPR001788; RNA_dep_RNAPol2.
 CC InterPro; IPR007095; RNA_pol_DS_PS.
 CC InterPro; IPR007094; RNA_pol_PSVit.
 CC InterPro; IPR002588; V_methyltransferase.
 CC InterPro; IPR000606; Viral_helicase1.
 CC Pfam; PF01443; Viral_helicase1.
 CC Pfam; PF01660; Vmethyltransferase; 1.
 CC Transferase; RNA-directed RNA polymerase; Helicase; ATP-binding.
 CC CHAIN 1 1616
 CC NP BIND 833 840 METHYLTRANSFERASE/RNA HELICASE.
 CC FT
 CC SEQUENCE 1616 AA; 183579 MW; AC7CF5429EAD49A CRC64;

Query Match 3.4%; Score 296.5; DB 1; Length 1616;
 Best Local Similarity 19.3%; Pred. No. 9.9e-10;
 Matches 303; Conservative 194; Mismatches 579; Indels 493; Gaps 63;

QY 24 QORAVKLPAPPLAKLETL--HRLYYPLRFKGGTLPRQHPI--LAGHQRVAE----- 72
 DB 42 RDRREKPVFSKVISSEQTLLATRAIPFQI--TYNNQNAVSHLAGLSLELEYLMQ 98
 QY 73 -----EVLHFA-----RGRSTVLEIGPSLSHSLKLGAPNAFVADYHGCTKYGTBDG 120
 DB 99 IYGSITVDIGGNFASHLPKGRAYV-----HCCM-----PNLDV-----KDI 135
 QY 121 SRHITALESRVAIGR-----DEPKADSLANGIASRTCCYDGVSCAFKSR----- 168
 DB 136 MRHEGQKSIIELYLSRLDRGKTVNPFQKAPDRYAEPEDAVCHNTQTCEHQPMQOSG 195
 QY 169 --VGIANSLVDVLTLELANAFENHGLHNVRAFMMMPBELLYMDVVAALGYRHVIEE 226
 DB 196 KYVALAHLSIYDIPADEFGAALLRNQVHTCYAHPSENLLEDDSYVND----- 245
 QY 227 PMAVNDCAFOGDDLEHPPELDFINESQERRIERIARAGSYRRAVITSGDDMDDAYLH 286
 DB 246 --EINACFSRQD-KLTF--SFASESTLNYCHFSNLIKVKCYKYPFASN--REVTMK 296
 QY 287 DF-----HTWL-----AYLLVR-----NYPPPFGLHIEVQR 314
 DB 297 EFLVTRVNTWFCFKSRIDTFLLYKGVANHSVSEQFYTAMEDAMHYKKTILANCSERILL 356
 QY 315 RHGSSIELRLTRAPRGCDMLAVNPTSGCLRNIFYYADASGTEHKTITLSQHKVNL 374
 DB 357 EDSSSVNWF--FKRMDVITLPLEDISL-----ETSKSRKREVLVSKQFVFV 402
 QY 375 LNFQTRPEKELVDVTLMSFARARLRAIVASVTESSWNISPADLVSTVSLVLTHT 434
 DB 403 LNHITTYQAKLTIANVL-SFVESIRSRVINGVTAAREMDVDKS-LLOSLSMTFFLH-- 458
 QY 435 ERRRAVAVKTAKQDV-----FGETS-----FWESLKHVLSGCCGLRLKGTDVVFTKRV 484
 DB 459 -----TKLAVLKDDLLISKFSLGKTVQGHVMDISLAFGN--APPSVK-----ERL 503
 QY 485 VDKYVHSLGDIICVRLSPBOVGLPSRVPPARV-FHDR----- 523
 DB 504 LNKKLIRAGDA-----LEIKPDLTYTFHDLVAEYKSSVDMALDIRKM 550
 QY 524 -----EELEYLRAGCVN-----E 537
 DB 551 EETEMVNAALSELSTVLRSDKDVDFVFSQMKSLSDVDPMTAKVIVAVMSNSGLTITFE 610
 QY 538 RP-----VPSTPVEEPO-----GPDAD----- 555

DB 611 RPTBANVAQLQDQEKASGALVYTSREVEPSMKSGMARGBLQLAGADHPRESSYSRN 670
 QY 556 -----LMH-ATAASLPEYR-----ATLQAGLNTDVKQOKI 584
 DB 671 BEISLSQFPAVATDSLIRKQMSIIVTGPILKQOMKPIFDSLVASLSAAVSNLVKILKD 730
 QY 585 TLBNALKT----- 592
 DB 731 TAAIDLETRQKFGVLDVASRKLILKPTAKSHAMGVETHAKRYVALLEYDEQGITCDD 790
 QY 593 -----IDGLTSPVRGL-----EMVSGPGSGKGTLLTALAESA 626
 DB 791 WRRVAVSESVYSDMAKLTIRLRILNDGEPHVSNAKVYLDVGPGCKTKYISLRVNP- 849
 QY 627 GSKALVYAPTRLEBANDR-----IKPSSATQHVALLIRBATAAGAPFATVVIDEC 681
 DB 850 -DEDLILVPGQAEMIRRRANSGLIYATKDNRVTDSFMNNPCKTRCQPKFLFIDG 908
 QY 682 FMPLVYVAIVHALSPSSRIVLVDVHQIGFID-----FQTSANMPLVADVKKCRRTF 737
 DB 909 LMLATGCVNPLVAMSLCDVAVYGDTOQIPIYINRVSGPPYPAHRSKLEVDVE--TRRT- 965
 QY 738 NOTKRCPADVYATTFPGSLYPG-CTTSGCYASISH-----VAPYRNSQAOGLCFT 788
 DB 966 --TLRCPADV--TRYLNRYEGFPVSTSVKKSVOEMVSGAAVNPISKLHGKILFT 1021
 QY 789 QEEKS---RHGAGAMTVHBAQGRTPFASVLIHYNGSTABQKLAEKSHLVGITRHTNL 845
 DB 1022 QSDKRALSRGYSBVHTVHVEQGETYSVLSRLTPTIISIAQDSPHVLAWSHTSL 1081
 QY 846 ----YIRPDGDIEROLNHSK-----AEVPTDIPAPLEITTVKPSBEQREVNATIP- 895
 DB 1082 KYTVVMDPLVSIIRDLKLSYLLDMYKVDAIGIQXQIDSV-----FKGSNLFVAAPK 1136
 QY 896 -----PGSATPHGAIHLNRKPGQPDCCGALAKTGYEYGRKAKIVE 940
 DB 1137 TGDISDMQFYDDKCLPGNST-----MNNPF-----DAVTRRLTDISLVNDCIILDMKS 1185
 QY 941 LAEPDAPKPRAPRGVQVQVKTNAHNGHQAOLTLISRYTKRSADLPLEHAKEDVKKML 1000
 DB 1186 VAAIPDQIKPL-----IPMRTAAMPRQGLLENVAMIKRNNAELSGIIDIENTA 1239
 QY 1001 N-SIDRHWDTWTBDAR-----DRAVEEOLKFTORG--GTVEDLBPDEYIRDI 1048
 DB 1240 SLVVDKFPDSYILKEKRPKNQVSLFSRESLNRLMEKQERVYIQGLADPFDVDAVQY 1299
 QY 1049 DFLAKTQ--QVSPRPITGKVGCGIAHNSLSLAPVLAAMIRILEIRL-----TSRT 1100
 DB 1300 RHMTKAQPKQLDTSIQTEYPALQTIIVYHASKKINAFGP--LFSELTRQLDSDVSSRF 1356
 QY 1101 VRYSGNLDEBEAMLEAKINOVPLATFVSAWTEPFDAHNHTSLPALLERIGTPAA 1160
 DB 1357 LFTFKRTAQIJBDFPGDD-SHP-NDVLELDISKIDYSQNEFHCAVYEIRRRIGFEDF 1414
 QY 1161 AVNLFRERCGRKTLRAKIG-SVEVDGLLDGAANTPCRNITPSAAVNLTFRGYKF--A 1217
 DB 1415 LGEVWKQGHRTKTLMDYTAIGKTCIMYQKSGDVTPIGNVTYIAACLSMLRMKELING 1474
 QY 1218 AFKGDLSL 1226
 DB 1475 AFQGDLSL 1483

RESULT 15
 RPO TMOB STANDARD; PRT; 1616 AA.
 AC P90211; Q83484;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE RNA-directed RNA polymerase (EC 2.7.7.46) (183 kDa protein) [Contains:
 Methyltransferase/RNA helicase (MT/HEL) (126 kDa protein)].


```

Qy 1189 --DGAAMTFCNRTIESAAMLTFRGVK--AAFKDDSL-----CGSHY 1231
Db 1447 ORKSGDVTTFIGNTVIAACMASMLPMEKVTKAFCGDDSLVYLPKCELPNIQCANLM 1506
Qy 1232 LRPDASRL-----HMGERYKTKHLKVEVOKIVPYIGILVSAEQVLDPVRSALKIFGRCY 1286
Db 1507 WNFPAKLFKKTYGYFCGRYVIHHDRAI-----VYVDPLKITSKLGAKHI 1551
Qy 1287 TSELVSKYVEAVRDITKGMDSARVHSLCHMSACYNTAPESAYIIDA VVRFGRGDP 1346
Db 1552 TDKEHLEEFRI SLADVSKSLNCAVYAQLDEAVREVKHTAP-PGSFVYKCI VKFLSNRVL 1610
Qy 1347 FEQL 1350
Db 1611 FESL 1614

```

Search completed: January 15, 2004, 16:50:01
Job time : 48.1935 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 16:06:47 ; Search time 68.131 Seconds

(without alignments)
6454.060 Million cell updates/sec

Title: US-09-991-262-40

Sequence: 1 MRAKATDVARYAAADVA...LKIRSTSDVDPVKISKA 1704

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL.23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_rvlnub:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	8671	98.2	1704	12	Q67724
2	1532.5	17.4	1925	12	Q9YRB3
3	540	6.1	1703	12	Q91116
4	536	6.1	1703	12	Q8A731
5	533	6.0	1707	12	Q8A731
6	529	6.0	1709	12	Q8A731
7	528	6.0	1708	12	Q8A731
8	525	5.9	1707	12	Q8A731
9	525	5.9	1709	12	Q8A731
10	522.5	5.9	1684	12	Q8A731
11	520	5.9	1707	12	Q8A731
12	520	5.9	1709	12	Q8A731
13	515	5.8	1698	12	Q8A731
14	512.5	5.8	1708	12	Q8A731
15	510.5	5.8	1708	12	Q8A731
16	503	5.7	1693	12	Q69410

17	499.5	5.7	1685	12	Q9E8G6
18	489	5.6	1693	12	Q89444
19	487	5.6	1693	12	Q8U7D6
20	494	5.6	1693	12	Q9W1L5
21	492	5.6	1693	12	Q81862
22	491	5.6	1693	12	Q81876
23	490	5.5	1693	12	Q81344
24	470	5.5	1693	12	Q9WC28
25	468	5.3	1693	12	Q8J734
26	467	5.3	1693	12	Q69418
27	466	5.3	1693	12	Q39221
28	406	4.6	727	12	Q81873
29	332	3.8	1616	12	Q81873
30	321.5	3.6	1596	12	Q84125
31	315.5	3.6	1616	12	Q911R9
32	314.5	3.6	1616	12	Q9JA04
33	308.5	3.5	1597	12	Q91PA8
34	305.5	3.5	1597	12	Q8V9A2
35	304.5	3.4	1597	12	Q8V9A6
36	302.5	3.4	1612	12	Q91B35
37	296	3.4	1612	12	Q10447
38	294	3.3	1694	12	Q05983
39	290	3.3	1116	12	Q091T0
40	284.5	3.2	1616	12	Q91BKS
41	284	3.2	1116	12	Q911R8
42	281.5	3.2	1616	12	Q9J943
43	279	3.2	1116	12	Q9JA03
44	273.5	3.1	1051	12	Q6108
45	269	3.0	631	12	Q913Y9

ALIGNMENTS

RESULT 1

Q67724 PRELIMINARY; PRT; 1704 AA.

AC Q67724;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE RNA-dependent RNA polymerase.
OS Helicoverpa armigera stunt virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
OC unclassified Tetraviridae.
OX NCBI_Taxid=37206;
RN [1]
RP SOURCE FROM N.A.
RX MEDLINE=21820274; PubMed=11831734;
RA Gordon K.H.J., Johnson K.N., Hanzlik T.N.;
RT "The larger genomic RNA of Helicoverpa stunt tetravirus encodes the
RT viral RNA polymerase and has a novel 3'-terminal tRNA-like
RT structure."
RT Virology 208:64-68 (1995).
RL EMBL: U18246; AAC98529.1;
DR EMBL; U18246; AAC98529.1;
DR InterPro; IPR001788; RNA_dep_RNApol2.
DR InterPro; IPR007095; RNA_pol_DS_PS.
DR InterPro; IPR006066; Viral_helicase1.
DR InterPro; IPR002588; V_methyltransf.
DR Pfam; PR00978; RNA_dep_RNApol2; 1.
DR Pfam; PR01443; Viral_helicase1; 1.
DR Pfam; PR01660; Vmethyltransf; 1.
DR PROSITE; PS50521; RDRP_VIRAL; 1.
DR RNA-directed RNA polymerase.
SQ SEQUENCE 1704 AA; 187841 MW; D7CPE086FB065997 CRC64;

Query Match 98.2%; Score 8671; DB 12; Length 1704;

Best Local Similarity 98.5%; Pred. No. 0;

Matches 1678; Conservative 2; Mismatches 24; Indels 0; Gaps 0;

QY 1 MRAKATDVARYAAADVAAYVLAQRAVAPPLKALSTLRLLYPLRFGKTLPPQ 60
DB 1 MRAKATDVARYAAADVAAYVLAQRAVAPPLKALSTLRLLYPLRFGKTLPPQ 60

QY 61 HPILAGHQRVAEEVLANFARGRSVLEIGSLHSALKLHGA PNA PVADYHGTCTKYGRDG 120
 DB 61 HPILAGHQRVAEEVLANFARGRSVLEIGSLHSALKLHGA PNA PVADYHGTCTKYGRDG 120
 QY 121 SRHTTALSSSVATGRPEPKADASLLANGIASRTTCYDGVSCAFKSRVIGIANSLYDVT 180
 DB 121 SRHTTALSSSVATGRPEPKADASLLANGIASRTTCYDGVSCAFKSRVIGIANSLYDVT 180
 QY 181 LEEELANAFENHGLHVAFMFMPEELL YMDNVNAEIGYRPHVIEEPMVAWDCAFQSGDL 240
 DB 181 LEEELANAFENHGLHVAFMFMPEELL YMDNVNAEIGYRPHVIEEPMVAWDCAFQSGDL 240
 QY 241 RLHPPELDFINESGERIERELAAAGSYSRRAVIFSGDDWDGDAYLHDFHTMALVLRNY 300
 DB 241 RLHPPELDFINESGERIERELAAAGSYSRRAVIFSGDDWDGDAYLHDFHTMALVLRNY 300
 QY 301 PTPPGFSLHIEVORRHSSIELRTTRAPPGDRLAVPRTSQCRIPIPIFYADASGTE 360
 DB 301 PTPPGFSLHIEVORRHSSIELRTTRAPPGDRLAVPRTSQCRIPIPIFYADASGTE 360
 QY 361 HKITLSQHKVNMILNFMOTRPEKELVDMTVLMSFARBLRAIVVASEVTESSNNISPAD 420
 DB 361 HKITLSQHKVNMILNFMOTRPEKELVDMTVLMSFARBLRAIVVASEVTESSNNISPAD 420
 QY 421 LVRIVSLVYLHIIEERRAAVAVKTAADVFGETSFMESLKHVLSGCCGLRNLKGTDFV 480
 DB 421 LVRIVSLVYLHIIEERRAAVAVKTAADVFGETSFMESLKHVLSGCCGLRNLKGTDFV 480
 QY 481 TKRIVDKRYVHSLDIIICDVRLSPOYGFPLBSRIPPARVFDRELEVLRAAGCYNRPV 540
 DB 481 TKRIVDKRYVHSLDIIICDVRLSPOYGFPLBSRIPPARVFDRELEVLRAAGCYNRPV 540
 QY 541 PSTPVEEPOGPDADLHMTAASLPEYRATLQAGLNDVKOLKTTLEVALKTIIDGLTSP 600
 DB 541 PSTPVEEPOGPDADLHMTAASLPEYRATLQAGLNDVKOLKTTLEVALKTIIDGLTSP 600
 QY 601 VRGLEMEYEGPPGSGKTGTLIALEBAAGKALYVA PTELEBAMDRRIKPPASATOHVAL 660
 DB 601 VRGLEMEYEGPPGSGKTGTLIALEBAAGKALYVA PTELEBAMDRRIKPPASATOHVAL 660
 QY 661 AILRRATAEGAPFATVVIDECFMPFLYVAIVHALSSPSRVLVGDVHQIGFIDPOGSA 720
 DB 661 AILRRATAEGAPFATVVIDECFMPFLYVAIVHALSSPSRVLVGDVHQIGFIDPOGSA 720
 QY 721 NMPLVRVVKOCRRRTNORTKRCPADVATTFEFOSLPGCTTSGCAVASHVAPDRNS 780
 DB 721 NMPLVRVVKOCRRRTNORTKRCPADVATTFEFOSLPGCTTSGCAVASHVAPDRNS 780
 QY 781 QAOTLCTOEKSHHGAEGAMTVHEAOGRTFASVILHYNGSTAEOKLAEKSHLLVGITR 840
 DB 781 QAOTLCTOEKSHHGAEGAMTVHEAOGRTFASVILHYNGSTAEOKLAEKSHLLVGITR 840
 QY 841 HTNHLVYRDPDPTGDIEROLNHSAKAEVFTDIPAPLEITTVKSEVORNEVMATIPPOSAT 900
 DB 841 HTNHLVYRDPDPTGDIEROLNHSAKAEVFTDIPAPLEITTVKSEVORNEVMATIPPOSAT 900
 QY 901 PHGAIHLIRKNFGOPGCGVALLAKTGEVGGRAKINVELAEADPAIPKPRAROEQVOW 960
 DB 901 PHGAIHLIRKNFGOPGCGVALLAKTGEVGGRAKINVELAEADPAIPKPRAROEQVOW 960
 QY 961 VKVTNASKHQALOTLLSRVTKRSADLPLHEAKEDVXKMLNSLDRHMDVTEBARPRAV 1020
 DB 961 VKVTNASKHQALOTLLSRVTKRSADLPLHEAKEDVXKMLNSLDRHMDVTEBARPRAV 1020
 QY 1021 FETOLKTOGCGTIEDLLEPDPYIRIDIPLMKTOQKVSPPVINTGKYGGCIIAHSKSLN 1080
 DB 1021 FETOLKTOGCGTIEDLLEPDPYIRIDIPLMKTOQKVSPPVINTGKYGGCIIAHSKSLN 1080
 QY 1081 FVLAAWIRILEEILRTSGRTVRSNGLPDEEAMMLEAKINOVHATFVSADWTEPDTAH 1140
 DB 1081 FVLAAWIRILEEILRTSGRTVRSNGLPDEEAMMLEAKINOVHATFVSADWTEPDTAH 1140

QY 1141 NNTSELLPALLERIGTPAAAVNLFRRCCKRTLRAXGLCSVEVDGLDGAAMTPCRRNT 1200
 DB 1141 NNTSELLPALLERIGTPAAAVNLFRRCCKRTLRAXGLCSVEVDGLDGAAMTPCRRNT 1200
 QY 1201 IFSAVWNLTLFRGCVKPAAFKGGDSILCGSHYTLFPDASRLHMGSRVYTKKHLYVQKIVPY 1260
 DB 1201 IFSAVWNLTLFRGCVKPAAFKGGDSILCGSHYTLFPDASRLHMGSRVYTKKHLYVQKIVPY 1260
 QY 1261 IGLVSAEQVVLDPVRSALKIFGRCTSELLSKYVAVNDITKMGSDARVHSLICMISA 1320
 DB 1261 IGLVSAEQVVLDPVRSALKIFGRCTSELLSKYVAVNDITKMGSDARVHSLICMISA 1320
 QY 1321 CYTNVAPESAAYIIDAVERFGKDPPEBQLRVVRAVQAPDAYSTTPANVRASCLDHVF 1380
 DB 1321 CYTNVAPESAAYIIDAVERFGKDPPEBQLRVVRAVQAPDAYSTTPANVRASCLDHVF 1380
 QY 1381 EPRQAAAPAGFVATCAKPEPSSLTAKAGVATSSHVATGTAPPESSWDAPANFSSEL 1440
 DB 1381 EPRQAAAPAGFVATCAKPEPSSLTAKAGVATSSHVATGTAPPESSWDAPANFSSEL 1440
 QY 1441 TPETPSTSSPSSSSSDSSTSCGRSLSGDPTARTTEDLNSRKPPSQDRSSSECLDRSG 1500
 DB 1441 TPETPSTSSPSSSSSDSSTSCGRSLSGDPTARTTEDLNSRKPPSQDRSSSECLDRSG 1500
 QY 1501 ERTGSSLTAPPADSPSSFSSEBARLATGPTVAAATSPSATPSCATDQVAARTTDPFAPFL 1560
 DB 1501 ERTGSSLTAPPADSPSSFSSEBARLATGPTVAAATSPSATPSCATDQVAARTTDPFAPFL 1560
 QY 1561 GSQSAKAVSKPYPPTTARKKAVTPLHAMGVGDRREVEDPETAIVQALLSGRPQK 1620
 DB 1561 GSQSAKAVSKPYPPTTARKKAVTPLHAMGVGDRREVEDPETAIVQALLSGRPQK 1620
 QY 1621 TKLSSDASKGYSRTKGSQSTSPAPSAADYQOARDCTVRCRAAEKARSCIEHPLASSA 1680
 DB 1621 TKLSSDASKGYSRTKGSQSTSPAPSAADYQOARDCTVRCRAAEKARSCIEHPLASSA 1680
 QY 1681 ASADLKIRISTSDSPDVKISKSA 1704
 DB 1681 ASADLKIRISTSDSPDVKISKSA 1704

RESULT 2
 QYR33
 ID QYR33 PRELIMINARY; PRT; 1925 AA.
 AC QYR33;
 DT 01-MAY-1999 (TREMBLrel. 10, Created)
 DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE RNA-dependent RNA polymerase.
 OS Nudauriella capensis beta virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
 OC Betatearavirus.
 OX NCBI_TaxID=85652;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=99263183; Pubmed=10329566;
 RA Gordon K.H., Williams M.R., Hendry D.A., Hanzlik T.N.;
 RT "Sequence of the genomic RNA of nudauriella beta virus (Tetreviridae)
 RT defines a novel virus genome organization.";
 RL Virology 258:42-53(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Gordon K.H.J., Williams M.R., Hendry D.A., Hanzlik T.N.;
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF102884; Acc97509.1; -
 DR InterPro: IPR001788; RNA_dep_RNAPol2.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR006066; Viral_helicase1.
 DR InterPro: IPR002588; V_methyltransferase.
 DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransferase; 1.
 DR PROSITE: PS50521; RDRP_VIRAL; 1.

RC STRAIN=JRAL;
RA Michiro S.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JRAL;
RX MEDLINE=21396683; PubMed=11504536;
RA Takahashi K., Iwata K., Watanabe N., Hatahara T., Ohta Y., Baba K.,
Mishiro S.;
RT "Full-genome nucleotide sequence of a hepatitis E virus strain that
may be indigenous to Japan."
RL Virology 287:9-12(2001).
DR EMBL: AP003430; BAB63939.1;
DR InterPro: IPR002589; A1pp.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR007095; RNA_pol_DS_PS.
DR InterPro: IPR007094; RNA_pol_PSVir.
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransferase.
DR Pfam: PF01661; A1pp; 1.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransferase; 1.
DR SMART: SM00506; A1pp; 1.
DR PROSITE: PS50507; HRP_POSITIVE; 1.
DR PROSITE: PS50521; KDRP_VIRAL; 1.
SQ SEQUENCE 1703 AA; 186780 MW; 8AB3C61B756967AE CRC64;
Query Match 6.1%; Score 540; DB 12; Length 1703;
Best Local Similarity 22.1%; Pred. No. 7.9e-27;
Matches 402; Conservative 176; Mismatches 560; Indels 678; Gaps 83;

13 AAADAVANVYVQBAVKLDLAPFAPKALET-LHRLYYP-LRFKGTLPPTQHPILAQH 68
22 AAASALANAVVVR-----PFLSRVQTEILINMQRQLVFPREV-L-WNHP1-----Q 68
69 RVAEEVLNFAFGS-TYLEIGPSLSALKLHGAPNAVADYHGC-TTYGTRDGSRHITA 126
69 RVHNELEYCRABAGRCLETGAHPRS--INDNENV--LHRCFLPVRGDVQRMVSA 121
127 LE-----SRSVATGRPEFKADASILANGIASRTFCVDGVGSCAFKRVGIANHSLYDV 179
122 PTRGPANCRSALRGLP-----VDRTYCFDGFSSCSFAEIVGALYSIHL 169
180 TLEELANAFENHGLMVAFMHMBELLVMDNVNAELGYRPHVIEEPMVAVKDCAFGCD 239
170 WPAVAAEMARHGMRTRYAALHLPRVELPPTG-----YHTSY----- 208
240 LRLHPELDFINESGERIERLAAGSYSRRAVIFSGDDMDGDAYLHFFHTMWLVL---- 295
209 LLIH-----DGNRAVVTYEGDTSAG--YNHVVSILRAWIRTK 244
296 LVNVPPTFGFSLHIEVQRHSGSIELRITRAP-----PGDEMLVVPRTSGLCRI 347
245 IVGHP-----LVIERVRAIGCHVLLITAAPESPMYPVPYRSTEVYRSITFGPGS 298
348 PNIFYVADASGTENKTIITLSQHKVNMILNFMQTRBEKELVDMTVLMSFAPARLAIIVAS 407
299 PSLF--PSACSTKSTFHAVPVHIMRLMLFGATLDDQAF-C-SRLMTYLRGISYKVTGA 355
408 EYTESNMWISPADLVRTVLSLVLIHIERRAAVN----- 443
356 LVANEGMASBDALAVITAAVLTITICHQRYLRTQAIISGMRLLEVEHAOKETRLYSWL 415
444 -KTADV-----FGETSEWES-----LKHVLGS--CCGIR 471
416 EKSQGDYVIRGQLQYVACCRRLSAGFHLDPRVIVFDESVPQRCTPLKXAGKCCCMR 475
472 -----NIKGTDVVFTKRVDKTVV- 490
476 WLGOECTFLPBAEGLVGDGHNEAVEGSEVDDQAEVPHLDVSGTYAVAHGQLEALYDAL 535
491 ---HSL-----GDIIIDVRLS-----PEQ--VGFL 510

536 NIPDIAARAAKLTAIVTELAAGPDLRCRTVLANKTFRITVTDGAILBANGPEQVLSFD 595
511 PSR-----VPPARV----- 519
596 ASRQMGAGSHSLTYELTPAGLQVKSISNGLDCAVPPGAPASAPRGVLAACSLALYR 655
520 --FHDRSEL-----EVLREAGCTNE----- 537
656 NRFTQHSILGGLMLHPEBGLGVPPPPSGHIESANPFCBGTLVTRWTSGSFSDFS 715
538 -----RPVPS-----TPPV-----SEPOG----- 551
716 PPEAABAPAAAGSSSPTPPVDIVLPPLSBPVGAPAPPTPEPARLPCCPTKNTP 775
552 -----FDAD-----LHMTAAVL 564
776 VRKPTAPPTERTRLTYTPDGAKYVAGSLFESDCMLVNASNPGHRPGGGLCHAFYQRF 835
565 PE-----YRAT----- 570
836 PEAFYPTPTMRBGLAAYTLTPRPIHNAVADYRVBQNPRLBAAYRTCSRRGTAAYPL 895
571 -----LQAGLMTDVQ-----LKITLENALKT--- 592
896 LSGGIVYPAGLSFDAMERNHRPQDGLYTEPAAAMEANKPTOPALTTI-EDTARTANL 954
593 ---ID-----GLTSPVRGLEMYE--GPPSGKTGTLTALAAAGKALYVAPT 636
955 ALBIDATDVGRACAGTISPP-GIVNYQTAGVPSGKRSI-----QGDDVVVVVPT 1007
637 RELREAMDRIKPPSASATQVNA-LAILRATBAGAPATTVVIDSCFMPPLVYVAIYVA 694
1008 RELSNWRRR--GFAFTPHTAARTIGRR-----VVIDEAPSLP-PHILLH- 1052
695 LSPSRIVLVGDVYQIQIDPFGGTSAMPLVRDVVQCRBRTTEQTRCADVATTFEQ 754
1053 MORASSVHLGDDPNQIADIDFENAGL-VPAIRP--ELAPTSWNVTRCPADV--CELIR 1107
755 SLVPGCTTSGCVASISHVAPDYRNSQAQTLCTOBEKSRHAGAMTVHAGQRTASV 814
1108 GAYPKIQTTGRVRLSLFWNER--AIGQGLVFTQAANKAN--PGALITVHAGQATTTET 1161
815 ILATNGSTABQKLAES--HLVGTIRHTNHLIYRPTGDIERQLNHS-----AK 863
1162 TII--ATAARAGLIQSSRAHAIVALTREHKECYILDPAG-LLRKVGISDVIYVNNPFIAG 1217
864 AEVFTDIPAPLEITTVPSSEVORN-EVMATIPQSATPHGAIHLKRNFGDDPCGVA 922
1218 GEVGHNR-----SVLPKGNPDNLDTLQAFPPSCQI--SAYHQLABELGHRP-APVNA 1268
923 LAKTGYEFGGRAXINVELAPDATPPKHPAQ-EGVQVWVVTNASKHQAQLTILSRYT 981
1269 VLPPCPLEBGLVMPQBLTVSDGV---LVFELTLDVHCMAAPSQRKAVLSTLVGRYG 1324
982 KRSADLPKLR-AKEDVYKMLNSLDRHMDWTYTBADRAVFEYQLKTKORCGTYEDLIEP 1040
1325 RKT--KLVEAHSADV--ESLARFTPTIGPVATTCELYELVEANWEKQDQDSAVLEL 1378
1041 D--DPIYRIDIDFLKKTQOK--VSPKIPNTGKGGGIIAHSKSLNVLVLAWIRIILE-BILRT 1096
1379 DLCNRDVSRIITFPQDCKKPTTGTITAHGKGGQISMSKTFCLFGPMFAIREKILAL 1438
1097 GSRTVRSNGLPDEEAMLLBAKINOVPHATFVSADMTBPTJANNNTSELLPAALLERIG 1156
1439 LPNIFVYG---DAVESVPAAAVSAGSAGVCFENDPSRFDSTQNNFSIGLECVVMEBCG 1494
1157 TPAAANVLFRRCGRKTLRA-----KGLGSVYVUGLDDSGAAMPCCRTTISAAVMTL- 1210
1495 MPQWLIRLYHIVRSAMITLQAKESLKGFWKXG-----SGRPTGLTMMTINMMIACHCY 1548
1211 -FRGVKPAAFKGDGS-LTCSHYLRFDASRLHMGERYTKILKYEVQKIVPYIGILVSAB 1268

```

Db      1549 EPRDFEVAAPFKGDDSVVLCSDYRQSRNAALINGCGIK---LKVDYRPIGLVAGVAVARG 1605
Qy      1269 QVVL-DPVRSALKI FPGCTSELYSKYVEAVDITKGSMDARYSHLLCHMSACYNYAP 1327
Db      1606 LGTLPPVVFPA---GR--LSE-----KNWGP-----GP 1628
Qy      1328 ESAAYIIDAIVVREGRG 1343
Db      1629 ERAEQLRLAVCDPLRG 1644

RESULT 4
08V731 PRELIMINARY; PRT; 1703 AA.
AC      08V731;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DT      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT      01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE      Nonstructural protein.
OS      Swine hepatitis E virus.
OC      Viruses; ssRNA positive-strand viruses, no DNA stage;
OC      Hepatitis E-like viruses.
OX      NCBI_TaxID=63421;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=swJ570;
RA      Okamoto H.;
RL      Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=swJ570;
RX      MEDLINE=21616952; PubMed=11741279;
RX      Okamoto H., Takahashi M., Nishizawa T., Fukai K., Muramatsu U.,
RA      Yoshikawa A.;
RT      Analysis of the complete genome of indigenous swine hepatitis E virus
RT      isolated in Japan.;
RL      Biochem. Biophys. Res. Commun. 289:929-936(2001).
DR      EMBL; AB073912; BAB79304.1; -.
DR      InterPro; IPR002589; A1PP.
DR      InterPro; IPR001788; RNA_dep_RNApol2.
DR      InterPro; IPR007095; RNA_pol_DS_PS.
DR      InterPro; IPR007094; RNA_pol_PS_Tir.
DR      InterPro; IPR000606; Viral_helicase.
DR      InterPro; IPR002588; V_methyltransfer.
DR      Pfam; PF01661; A1PP; 1.
DR      Pfam; PF00978; RNA_dep_RNApol2; 1.
DR      Pfam; PF01443; Viral_helicase1; 1.
DR      Pfam; PF01660; Vmethyltransfer; 1.
DR      SMART; SM00506; A1PP; 1.
DR      PROSITE; PS50507; RDRP_POSITIVE; 1.
DR      PROSITE; PS50521; RDRP_VIRAL; 1.
SQ      SEQUENCE 1703 AA; 187070 MW; 518FCB12851FA675 CRC64;

Query Match      6.1%; Score 536; DB 12; Length 1703;
Best Local Similarity 21.6%; Pred. No. 1.5e-26;
Matches 333; Conservative 183; Mismatches 558; Indels 686; Gaps 81;

```

```

Qy      240 LRLHPHELDFINESQERRIERLAAGSYRRAVIFSGDDMDQDAYLHDFHTLALYL----- 295
Db      216 -----RAVVTYEGDTSAG--YHNDVSIILAWIRTK 244
Qy      296 LVANVPTPFGSLHIEVORRHGSSIELRITRAP-----PDRMLAVPRTSOGICRI 347
Db      245 IYGDHP-----LVIERVAIGCHPULLTAAPBEPSPMYVPRFRTETVYVASIFPGGS 298
Qy      348 PNI FYVYADASGTEHKITLTSQHKVNMLNFMQTRPEKELVDMTVLMSFARARLAIVVAS 407
Db      299 PSLF--PSACSTKSTFPAVPHVIMDRMLFGATLDQAFRC--CSRLMTYLRGISYKVTGVA 355
Qy      408 EYTESMNISPADLVARTVSVLYLHIERRAAVV----- 443
Db      356 LVANEGMNASBDALTAVITAAVLTTCORLYRTQALSKGRRLVEHAKQFTRLYSWLP 415
Qy      444 -KTAQDV-----FGRTSPFWS-----LHNVIGS--CCGUR 471
Db      416 EKSGRDYIFGRQLQFYACRRMLSGAFHLDPRVLVFDBSVPCRCRTFLKKVAGKFCCEFK 475
Qy      472 -----NLKGTDVFTKRVVDKYRV- 490
Db      476 WLQGBCTFLPABAGLVGDQGHNDNAYEGSVDPAPAPAFVDSGTVAVHGRLBALRYAL 535
Qy      491 ----- 490
Db      536 NIPHDIAABARLTATVBLAASPDRLCRTIIGNKTFTTVVDGAILVNGPEQVYLSPD 595
Qy      491 -----HSL-----GDI--ICD----- 499
Db      596 ASRQTMGAGSHSLTTELTAGLOVRISNGLDCTAVPPGAGAPSAAPGEVAACSLARY 655
Qy      500 -----VRLSPQV-----GF----- 509
Db      656 NRTQSHSLTGLMHPBGLIGIFPPSPGHIMESANPPCGSGTLTRTMTSGFSFSDPS 715
Qy      510 -----LBSVPPARVFDREBELYLKAGCTNBRPVSTP----- 544
Db      716 PPEADHAPALITSGLPHSSTPPA--SDIWLPPSPSBSQVDMAFPVAPKPVGLPESNBP 772
Qy      545 --PVEEPQ-----FQADL--W----- 557
Db      773 ITPVRKPPAPASRTKRLLYTPPDGAKVYAGSLFSDCMLVNASPGRPGGLCHAFY 832
Qy      558 -----HATTAASLPERYA-----TLQAG----- 574
Db      833 QREPAFYPTPIRMREGLAAYLTTPRPIIHAVA--PDTRVGNPRRLBAARETCSRRG 869
Qy      575 -----LMTDVQQLKTL-----ENAL 590
Db      890 TAAVPLGSGIYQVPVSLFDMERNHRPGDELYLTERPAAMPBANKPAQPALTTIEDTA 949
Qy      591 KT-----ID-----GLTSPVRLGLEYB--GPPGSKCTTLTALAEAGKAL 631
Db      950 KTNALALEIDANTEGRACAGCTIRP--GVNHQFAGVPGSKSRST-----QQGDVDV 1002
Qy      632 YVAPTRLEBANDRIKPPSASATQVALAIL--RRATAGAPAFATVVIDCEMFPLVVY 689
Db      1003 VVPTTELANSMRR--GPAAPTHTTAARVTTGR-----VVIDEAPSLP--PHL 1048
Qy      690 AIVHALSPSRIYVGDVHQIGFIDFGSTANPLVADVVKQCRRTFNQTKCPADVVA 749
Db      1049 LILH--MQRASSVYLLCDPNQIPALIDEHAGL--VPAIRP--ELAPTSMMWHYTHRCPADV-- 1102
Qy      750 TTFPGSLYGCCTTSGCVASISHVAPDYNSQAQCTCFPOEBSRRGAGAGAMVYHQAQR 809
Db      1103 CELIRGATPKIQTSSVLSLFTNNEP--AGQKLVFOAAAN--PGATVHBAQGA 1156
Qy      810 TPAVSLIAYNGSTAOKLAERK--HLVAGTRHTNHLIYRPTGDIERQLNHS----- 861
Db      1157 TTFETTTI--ATADARGLIQSBRHAIVALTNRHTEKCIILDRPG--LAEVGISDVIYVN 1212
Qy      862 ---AKAEVFTDIPAPLEITTVKPSSEVQNRNEMATTIPQSAITPDHGAIHILRNKFGQDPDC 918

```

```

Db      1213 FFLAGGEVGNHRRS--VLRPNPDONT--ETLQAFPPSCOI--SAHQLEELGHRP-A 1264
Qy      919 GCVAAKTGYEVFGGAKINVELABPDATPKHRAFC-EGVWYKVTNASHKQALQTL 977
Db      1265 PVAALVPPCPBELGGLYMPQLVSDSV-----LVFELTDLVHGRMAAPSRKAVLSTLV 1320
Qy      978 SRYTKRSADLPLEH-AKEDYKRMNSLDHMDWTVEBDARDRAVETQLKTRQGTVED 1036
Db      1321 GRYGRRRT---KLVEAASHSDVR---ESLARFIPTIGPVATTCCLYELVEANVEKQDQSA 1374
Qy      1037 LLEED--DPYIRNDIDELMKTOOK-VSPKINTGVGCGIATASHKSLNPLVLAAMRIIE-E 1092
Db      1375 VLELDLCRDVSRITTFQKDKCNKFTTGETIAHGRVGGISAMSKTFCALFGPWFRAI EKE 1434
Qy      1093 ILRTGSTRVRSNGLPDEEEMLEAKINQVPHATPVASDWTPEPTAANNTESELLFALL 1152
Db      1435 ILALLPPIFYG----DAYEGSVFAAANVSAGSCMVFENDSFEDSTQNNFSLGLECVVM 1430
Qy      1153 ERIGTPAAVNLFPERCGRKTLRA----KGLGSVEVDGLIDSGAAMTPPCNNTIFSAAVM 1207
Db      1491 EECMPQWLRLYLHVSAMILQAPKESLKGFWKKH-----SGEPGTLINNTVMNMAII 1544
Qy      1208 LTL--FRGVKFAAFKGDG--LLCGSHYLRFDASRLHMGERYKTHLKVEQKIVPIGLT 1264
Db      1545 AHCFEPDLRVAAFRKGDSDSVLCSDFQSRMAAALACGK--LKVDYRPIGLYAGVV 1601
Qy      1265 VSABQVVL-DEVRSAALKIFGRCTSELLSKYVEAVRDITKGWSDARVHSLCHMSACY 1323
Db      1602 VAPGLGTLPDVVRFA---GR--LSE-----KNMG------ 1626
Qy      1324 NYAPESAAYITDAVVRFGRG 1343
Db      1627 --GERAEQLRLAVCDFLRG 1644

RESULT 5
Q8JUN7 PRELIMINARY; PRT; 1707 AA.
AC Q8JUN7;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE ORF1 protein.
OS Hepatitis E virus.
OC Viruses; sRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAK-Sai;
RA Mishiho S.;
RL Submitted (NOV-2001) to the EMBL/Genbank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JAK-Sai;
RA MEDLINE=2196248; PubMed=12001054;
RA Takahashi K., Kang J.-H., Ohnishi S., Hino K., Mishiho S.;
RT "Genetic Heterogeneity of Hepatitis E Virus Recovered from Japanese
RT Patients with Acute Sporadic Hepatitis.";
RT J. Infect. Dis. 185:1342-1345(2002).
RL EMBL; AB074915; BAB93536.2; -.
DR InterPro; IPR002589; Alpp.
DR InterPro; IPR001788; RNA_dep_RNAPol2.
DR InterPro; IPR007095; RNA_pol_DS_B9.
DR InterPro; IPR007094; RNA_pol_PSVir.
DR InterPro; IPR006066; Viral_helicase1.
DR InterPro; IPR002588; V_methyltransf.
DR Pfam; PF00978; RNA_dep_RNAPol2; 1.
DR Pfam; PF01443; Viral_helicase1; 1.
DR Pfam; PF01660; Vmethyltransf; 1.
DR SMART; SM00506; Alpp; 1.
DR PROSITE; PS05057; RDRP_POSITIVE; 1.

```

```

DR PROSITE; PS05057; RDRP_VIRAL; 1.
SQ SEQUENCE 1707 AA; 186724 MW; CSEF8DDEBC9313B9 CRC64;
Query Match 6.0%; Score 533; DB 12; Length 1707;
Best Local Similarity 21.5%; Pred. No. 2,4e-26;
Matches 376; Conservative 171; Mismatches 567; Indels 620; Gaps 70;
Qy      2 YAKTDAKVVAAADVAVANVLOQRAVKLDPAPELKALET--LHRLYTP--LAPKGTLP 57
Db      6 FIKAPGVTTAIEQALAAANALANAVV--RPELSRLQREILINLQRPQLVRFVTL- 62
Qy      58 PTOHPILAGORVAVETLHNPARGS--TVLEIGSLSALKTHAPAPVADYHGC--TKY 115
Db      63 -WNHPI---QRYLHNELEQYCARAGRCLEVGAPRS--INDPNV---LHRCPLKP 110
Qy      116 GTRDSRHITALE-----SRSVATGRPEFKADASTLANGIASRTFCVGVSCAKSR 168
Db      111 VGRDVQWYTHAPTRGPANCRRLRGLP-----ADRTYCFDGFSGCTPAAE 158
Qy      169 VGIANSLVYVTEBELANAFENHGLHVRAPMGMPEELLYMDVNVNAELGYRFFVIEBP 228
Db      159 TGVALYSILHDLMPADVABAMARHGMTRLYALHLPEVLLPFGYHTSYLLIH----- 212
Qy      229 AVNOCARQGGDLRIHFPELDPIFNSQERRIBRLAARGSYGRRAVIFGDDDKGDAYLHDF 288
Db      213 -----DGD-----RAVITYGDSAG--YNHDV 233
Qy      289 HTWLAVLVNRYTPFGPSLHIEVQRHGSIEILRTAP-----PGDRMLAVVPR 340
Db      234 SILBAW--IKTKTGPHPLVIERVAVGCHFVLLTAAPEBPMFVYPRSTEVYRS 291
Qy      341 SGLCRIPNIFTYADASGTEHKTYLSQHKVNMILNFMQTRPEKELVDMVLMSPADRL 400
Db      292 IFPGGSPSLP--PSACSTKSTFAVAVPHIMDRILMCGATLDDQAF-CSTLMTYLGIS 348
Qy      401 RAIYVASEVTSSWNISPADLVRTVSVLYLHIIERRBAVAV----- 443
Db      349 YKTVGALVANEGNASEDALTAIVITAYLITGHQRYLRTQALSKGMKRLLEHNAQFIT 408
Qy      444 -----KXADQVFG-ETSPWESLKHVLS----- 466
Db      409 RLVSMLPEKSGRDYIPRQLQFYAQGRWLSAGPHLDPRVLVPDEAAPCRGRSLLRKAH 468
Qy      467 --CCGLANL-----KGTDV----- 478
Db      469 KFCFPMWLGQDCTFLQPIEGRVGQYDNEAFEGSDVDPABEATVISIGSYLVTSQQL 528
Qy      479 -----VFTK----- 482
Db      529 QPLYQALGISDULAARAGRLTATVGVSDADGRLTCTIMGNKTTFTVFTDGAQLVNGPE 588
Qy      483 -----RVYDK----- 487
Db      589 QYVLSFDLAKQTMAGPHSLSYVLTSAGLBRVVYSAGLDCAVPSGVATPSAGVSAF 648
Qy      488 ---YRV-----HSL----- 493
Db      649 CSALYRFRKCVQSRSLTGLMYPEGLIGLPPPAFGHTWSTNPFQGESLTYTRTWSVS 708
Qy      494 GDIIICDVRLSPQVGLPSRV--PPRVFPHRBEELVLRBAQCYNBRVPSTP----- 544
Db      709 GFSSCFBFLBPASGSLPPRAIDPVTVDAPSPSILALPRSVBQTTPLDPAQDAAS 768
Qy      545 -----PVBPQ-----FDAD----- 555
Db      769 APPGAPVPAPAPRVTHPSGPRRLHTYVDGSKVYAGSLFBESECTLVNANSPGHRG 828
Qy      556 --LHATRAASLP-----YRAT 570
Db      829 GGLCHAFYQRPESFDPABFWSGDFAAYTLTPRPIIHAAVDYRVEHNPRLAAAYRET 888
Qy      571 -----LQGLNTD----- 578

```


Db 889 CSRRGTAAYPLGAGIYKVYVGLSPDAMERNHRPGDELVTBPAIAFMFANRPTLPALATI 948
 Qy 579 -----VKOLKITLENAL---KTIIDGILTSFVNGLEMYE---GPPSGKGTGTLIAALBEA 626
 Db 949 TEDTARTANIALALEDSATEVGRACACRVER--GVNHYOPTAGVPSPGSKRSI-----QQ 1001
 Qy 627 GKKALVATPBLREBAMDRIRKPPSASATQHVALLATIRATAGAPATVVIDECMFPL 686
 Db 1002 GEVDVVVVTRELRNSWRRR---GPAAYTPHTAARTRG-----RAVIDEALAP 1049
 Qy 687 VYVAIVHALSPSSRIYLVGDVHOGIFIDFQGSANMPLVDVKCRRTFNOTKESCPAD 746
 Db 1050 PHLLILH--MQRASSVHLGDPNQPIDAFDEFHAGL--VPAIRP--ELVPTKMHILTHRCPAD 1105
 Qy 747 VVATTFQSLYRQCTTSGCVASISHVAPDYRNSOQTLCTQEEKSRHGAEGAMVYHEA 806
 Db 1106 V--CELRIGAYPKIQASRLRLSLFWGEPRV---CQKVFQTAALAAAN--PGAITVHEA 1157
 Qy 807 QGRTFASVILHYNGSTAEOKLAEKS--HLVIGITRHTNLVIRDPDGIEROLNLSAKA 864
 Db 1158 QGATFTEITII--ATADARGLIQSSRAHAIYALHTHTKCVVVDAPG--LLREVGIS--DA 1212
 Qy 865 EVFTDIPABLETITVVRSEVORNEVWATIPQSAAP---NGAHLNRKNGDQDCG 920
 Db 1213 IVNNFPLSGGQIGQHRS--VIRGTVDNSVDTLDAPRPSQGFSAHYQLABELGHRP-AP 1270
 Qy 921 VALAKTGYEVFGGRAKINVELAEPAATPKPHAPQ--BGVQWVKVTNASKHQAQLQTLR 979
 Db 1271 AAVLPCEPELEOGLLYMPOELTTSDSV---LTFEITDIYHCHMAAPSQKAVLSTLVGR 1326
 Qy 980 YTKRSADLPHEAKE--DYKMLNSLDHMDWTETEDARDAVFEQTLKTFQSGTVEIDL 1038
 Db 1327 YGRRT--KLYEAAHADVKGSLNHFTPELGPISTVTTCELYELVEMAEKQDSAVLEL- 1382
 Qy 1039 EPDDPYTRD---IDFLMKTQOK--VSPKPIITGVGOGIAHNSKSLNFWLAAMIRILE-B 1093
 Db 1383 ---DLSRVSRVSRITTFQKONKFTTGEBTIAHGKVGOGISMSKTFCLPFPWRARIBKE 1439
 Qy 1094 LRTGRTVRYNSGLPDEBEAMLEAKINOVPHATFVSADMTERTDTHNTSELLPAALLE 1153
 Db 1440 LAAALANVFEFG---DAYEDTVLAAVAGAPGCKVEFENDSEFDSQNNFSLGECIIME 1495
 Qy 1154 RIGTPAAANLPR-----ERCGKRTIRAKGIGSVFVGLDLSGAAMPFCRMTTISAAY 1206
 Db 1496 ECGMPQWMTLHYLVASAMVLOAPKESLR--GFWKCH-----SGBERTLMTVMMAV 1547
 Qy 1207 MLTL--FRGVKFAAFKGDSD--LLCGSHYLRFDSARLHMGERYTKHLKVEQKIVPYIGL 1263
 Db 1548 IAHCEYFRDLKVAAAFKGDSDSVLCSIDYRGRNMAALITGCGLK---LKVDFRPIGLYAGV 1604
 Qy 1264 LVSAEQVL--DPVRS 1278
 Db 1605 VVAPGLGTLPDVVRFA 1620
 RESULT 6
 Q8JUM6 PRELIMINARY; PRT; 1709 AA.
 ID Q8JUM6
 AC 08JUM6
 DT 01-OCT-2002 (TEMBLrel. 22, Created)
 DT 01-OCT-2002 (TEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE ORF1 proteinh.
 OS Hepatitis B virus.
 OC Viruses; sRNA positive-strand viruses, no DNA stage;
 OC Hepatitis B-like viruses.
 OX NCBI_TaxID=12461;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=JMY-Haw;
 RA Mishiro S.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 [2]

RP SEQUENCE FROM N.A.
 RC STRAIN=JMY-Haw;
 RX MEDLINE=21996248; PubMed=12001054;
 RA Takahashi K., Kang J.-H., Onihimi S., Hino K., Mishiro S.;
 RT "Genetic Heterogeneity of Hepatitis B Virus Recovered from Japanese
 RT Patients with Acute Sporadic Hepatitis.";
 RL J. Infect. Dis. 185:1342-1345 (2002).
 DR EMBL: AB074920; BAB93541.2; --
 DR InterPro: IPR002589; AIDP.
 DR InterPro: IPR001788; RNA_dep_RNApol2.
 DR InterPro: IPR007095; RNA_pol_DS_PS.
 DR InterPro: IPR007094; RNA_pol_PSYR.
 DR InterPro: IPR006060; Viral_helicase1.
 DR InterPro: IPR002588; V_methyltransfer.
 DR Pfam: PF00978; RNA_dep_RNApol2; 1.
 DR Pfam: PF01443; Viral_helicase1; 1.
 DR Pfam: PF01660; Vmethyltransfer; 1.
 DR SMART: SM00506; AIDP; 1.
 DR PROSITE: PS50507; RDRP_POSITIVE; 1.
 DR PROSITE: PS50521; RDRP_VIRAL; 1.
 SQ SEQUENCE 1709 AA; 187490 MW; 6878210D64307A24 CRC64;
 Query Match 6.0%; Score 529; DB 12; Length 1709;
 Best local Similarity 21.6%; Pred. No. 4,4e-26;
 Matches 396; Conservative 182; Mismatches 552; Indels 700; Gaps 83;
 Qy 13 AAADVAAYANVLOQRAVKLDPAPPLKALET--LHLRYTP--LAFKGTLPPTOHPILAGHQ 68
 Db 22 AAANSALANAAYVR-----PFLSRVQTEILINLMQPRQVLVRPREVL--MNNPI-----Q 68
 Qy 69 RVAREVLNHPARGS--TVLEIGPSLHSAKIHGAAPAAVDYNGC--TKYGTGSGSHITA 126
 Db 69 RVTHNELBOYCRRARCRCLVEGNHPRS---INNNPV---LHRCFLRPVGRVQRYSA 121
 Qy 127 LR-----SRSVATGRPEFKADASILANGIASRTCCVDVGSCAFKSRGVIANHSYDV 179
 Db 122 PIRGPANCRSALRLRP-----ADRTYCFQPSRSCAFALRGVALYSLHDL 169
 Qy 180 TLEELANAFENHGLHVRAPMHPMBELLYMDNVNNAELGRRFVIEBPAAVKDCARQGD 239
 Db 170 WPAADVAAARAHQMTELYVNLPRVLLPRTGYHTTSYLLIH-----DGD 215
 Qy 240 LRHPFELDPFINSQSRRIERLAAAGSYGRBAYIFSGDDDWGAYLHDFHTMLAYL----- 235
 Db 216 -----RAVTTYGDYSAG--YNNDSILRAWLRTRYK 244
 Qy 296 LVANVPYTPFGFSIHIEVORHSGSIELRTTRAP-----PQDRMLAAVPRTSQGLCRI 347
 Db 245 IYGDHP-----LVIERVALIGCHFYLLTAAPBSPMYPVPRSTEVYVRSIRPGGS 238
 Qy 348 PNIIFYADASGTBHKITLTSQHKVNMLNFMQTRPEKELVDMTVLMSFAPARLRAIVVAS 407
 Db 299 PSLP--PSACSTKSTGFAHVNIIMDRMLFGATLDQDAFC--CSRLMTYLRGISYKVTGVA 355
 Qy 408 EYTESSWNISPADLVATVTVSLYVLIIEERRAAVAV----- 443
 Db 356 LVANEGMNASBDALTAIVITAAVLTICHQVRLTAISKGRRLVEBHAOKPTIRLYSWLF 415
 Qy 444 -KTAQDV-----FGTSIFMES-----LKHVLGS--CGGLR 471
 Db 416 EKSGRYITGROLOFVACGRMLSAQFHLDPRLVVDSEVPCRCRTFLKVAAGKPCCFMR 475
 Qy 472 NL-----KSTDV----- 478
 Db 476 WLQOECTFLPAGLVGQGHNDNAVBSGVDPAPRAHLDVSGTYAVHGHQLEALYRAL 535
 Qy 479 -----VPTKYVDKYAV----- 450
 Db 536 NVPHDIAABASRLTATVELVASPDRLECRVLGKNTFRITVVDGAHLBANGPQYVLSFD 595
 Qy 491 -----HSL-----GDI--ICD----- 499
 Db 596 ASRQSMGAGSHSLTYELTRAGLQVRISSNGLDCTAVFPPGAGAPAAFGVAAACSAHYRY 655

```

Qy 500 -----VRLSPQV-----GF----- 509
Db 656 NRFTORHSLTGLTGLHPEGLGIIPPPSPGHIMESANPCEGLTYRTWTSGSSDPS 715
Qy 510 -----LPSRVPARVHDELEVL---REAGCYMRPVSTP----- 544
Db 716 PPEAAAPVPAAPGLPHPTPV-----SDIWLPPESGQIDAPVPVKTIVGLPSP 769
Qy 545 -----PVEPQO-----FDADL-W----- 557
Db 770 IVLAPPPPPSPVKKSPSPSRIRRLIYTPDQARVYAGSLFESDCMVLNANSPGHRP 829
Qy 558 -----HATPALSPEYRA-----TLQAG 574
Db 830 GGGGLCHAFYQRPPEAFYPTTEFIMEGLAAYTLTPRIIHAVA---PDYRVEQNPRLEAA 886
Qy 575 -----LMTDVQOLKITL----- 586
Db 887 YRETCSRRTAAYPLLSGCIYQVPVLSFDAMERNHRGDELYLTPAANFEANKSPQ 946
Qy 587 -----ENALKT-----ID-----GLTSPVRGLMEYR---GPPSGKTGTLIA 622
Db 947 ALTTEDTARTANALAEIDATVEGRACAGCTISP--GIYHQFAGVPGSGKRSI--- 1001
Qy 623 LEAAGKALYAPFTRLEBAMDRIKPPSASATQVA--LAILRATMEGAPFATVIDE 680
Db 1002 --OOGDDVVVVPPRELNSWRR--GFAAFTPTAARVITGR-----VVIDE 1046
Qy 681 CFMPPLVVVAIVHALSPSRIVLVVDVHIGRIDQGSAMPLVRDVVQCRRTFTQ 740
Db 1047 APSLP-PHLLIHL-MQRASSVHLADBNQIPADIDENAGL-VPAIRP--ELAPTSWMT 1101
Qy 741 KRCPADVATTFQSLYPCGCTTSGCVASISHVADYNSQAOITCPTOEKSRHAGBA 800
Db 1102 HRCPADV--CELIRAYPKIQTTSRVLSLFMNEP-----ALGQKLVFOAKAAN--PGA 1153
Qy 801 MTVEAOCRTASVYLHNGSTAEOKLIAEKS--HLVGTIRHNLHYRPTGIEIOL 858
Db 1154 ITVEAOCATFETETII--ATADARGLIOSSRAHAIYALTFTTEKCYILDAFG--LIREV 1209
Qy 859 NHS-----AKAEVFTDIPAPLEITTVKSEVORN--EVMATIPQSAATPHGAIHL 908
Db 1210 GISDVIYNNFLAGEVGHHR-----SVIPRGNPDQGLTQAFPPSCQI--SAHQOL 1261
Qy 909 RKNFGDQPCGCVALAKTGYEVFGRAKINVELAEPDAPKPRHAFQ--BGVQWVYVNAS 967
Db 1262 ABEIGHRP-APVAAVLPCEPLEGGLVMPQELTVSDSV--LVFELTDIYHGMAPS 1316
Qy 968 NKQALQTLRSRYTKRSADLPHE-AKEDYKMLNSLDHMDWTTEDEARAVETOLK 1026
Db 1317 QKAVLSTLVGRYGRRT--KLYEAAHSIVR--ESLARFPTIGPVQATTCLELVEBA 1370
Qy 1027 FTORGATVEDLEPP--DPYIRIDIDFLMKTOOK--VSPKIRTKGKGGAHSHKSLNVL 1083
Db 1371 MVEKQDSAVLELDLCNRDVSRTTFQKQCNKFTGTETIAHGKQGGISAKSKFCAIF 1430
Qy 1084 AAMIRILE-ELITGSRTRYNSGLPDEBEAMLEAKINQVPHAFVSAWMTPEPTAHNN 1142
Db 1431 GMPFAIETKEILALPPIFYG---DAVEESVFAAASGAGSCNVFENDPEFSTQNN 1486
Qy 1143 TSELLFALLERITGPAAAVNLPRRCGRKTLRA-----KGLGSVEVDGLDLSGALWPC 1197
Db 1487 FSLGECVMECEGMPOMLIRLYHLVRSAMIIQAPKESLKGPMKX-----SGEPGTL 1540
Qy 1198 RRTISAAMVLTU--FRGVKFAAFKGDG--LTCGSHYLRFDASRLHMGERYTKLKYEV 1254
Db 1541 MNTVMNMAIIAHCFRDRVAAFKGDDSVLCSIDYRGRNAALIAAGGLK---LKDY 1597
Qy 1255 OKTIVEYIGLVASAEQVLT--DPVRSALKIFGRCTYSELLSKYVEAVRDTTKWMSDARYHS 1313
Db 1598 RPIGLYAGVVAVPGIPLDVVRFPA---GR--LSB-----KNWGP----- 1632

```

```

Qy 1314 LICHMSACTYNNVAPESAAYIIDAIVRRG 1343
Db 1633 -----GPRABQRLALVCDPLRG 1650

RESULT 7
ID Q8BB07 PRELIMINARY, PRT: 1708 AA.
AC Q8BB07
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE RNA polymerase.
OS Swine hepatitis B virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis B-like viruses.
OX NCBI_TaxID=63421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkell;
RX MEDLINE=22297120; PubMed=12409369;
RA Pel Y., Yoo D.;
RT "Genetic Characterization and Sequence Heterogeneity of a Canadian
RT Isolate of Swine Hepatitis B Virus."
RL J. Clin. Microbiol. 40:4021-4029(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Arkell;
RA Yoo D., Pel Y.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY15488; AA092182.1;
SQ SEQUENCE 1708 AA, 187805 MW, 3379DADA519EA2BF CRC64;

Query Match 6.0%; Score 528; DB 12; Length 1708;
Best Local Similarity 21.4%; Pred. No. 5,1e-26;
Matches 379; Conservative 177; Mismatches 542; Indels 670; Gaps 76;

13 AAADVAVANVQCAVADFPAPKALET--LHLLYV--LRFGGTLPPYOHPILAGH 68
22 AAANSALANAVVR-----PFLSRVQTEILIMQROLVFPRPV--WNHP-----Q 68
69 RVAESVLNFPARGS-TVLEIGPSLSALKHGAHPNAVADYHGC--TKYGRDGSRIHTA 126
69 RVHNELEQYCRABAGRLGVGAMPRS--INDPNV---LHRCFLKPVGRDVORWTSA 121
127 LE-----SHSVATGRBEFPAADSLANGIASRTFCYDVGSCAFKSRVGIANHSLYDV 179
122 PTPGPAANCBSALRGLEPP-----VDRTYCFDGFSCRCAFANKGVALYSIADL 169
180 TLBELANAFENHGLHMPVAFHMPBELLYMNVNNAELGYTFHYVIEBMAVYKDCAFQGGD 239
170 WPAVDAEMARHGRTLRALALHLPPEVLLPQGT-----YHTTSY----- 208
240 LRLHPELDPIINSGERRIBRLAAGSYRRAVIFSGDDWDGADYLDHFTMLAYL----- 295
209 LLIH-----DGSRAVVTYEGTSGAG--YNHVVSILKAMIRTK 244
236 LVNRYPTPFGSLHIEVGRHSGSIELRITAP-----PGDRMLAVPRTSGLCRI 347
245 IVGDHP-----LVIERVRAIGHFVLLTLTAPEBSPMPVYPYPRSTEVYVRSIPGGS 298
348 ENIFPYADASGTEHKITITSGHKVMNLINPQGTREKELVDMTVLMSAPARLAIYVAS 407
239 PSLF--PSACGKSTFPAVPHVINDRLMLFGATLDDQAF-C-CSRIMTYLRGISYKVTGA 355
408 EYTESNMISPADLVRTVVSILYVLIHIERRAAVAN----- 443
356 LVANEGMAASDALTAIVTAIVLITCHORYLRTQALISGMRLEVBHAKPTRLYSWLF 415
444 -KTADV-----FGSTFWES-----LKAVLGS--CGGLR 471
416 EKSGRDVIQRQOLFVACRRWLSAGFHLDPRLVLFDESVPSCRCTPIKTVAGKPCCFMK 475

```

```

Qy 472 -----NLKQDVVFKRVVKKRV- 490
Db 476 WIGQECTFLERABGLVGDGHNEAVEDEVDRAERAYLDVSGTAVYGHQLEALYRAL 535
Qy 491 ----- 490
Db 536 NIPHDIVARASRLTAVELVASPRLECRITVLGNKVFRTTVDCAHLEVNGPBOHVLFD 595
Qy 491 -----GDIICDVLSPBOVGLPSRV- 514
Db 596 VLRQSMGAGSHSLTYELTSLAGLOVRISNGIDCTATSP--GGAPAPABGEVAAFCSAL 652
Qy 515 -----PPARVFHRELEVLREAGCYNER- 538
Db 653 YRNRFTQRHSLTGLMLHPEGLLGITFPSPGHIMESANPFGEGTLYTRTSTSGFSS 712
Qy 539 -----PVPSTPPVE- 548
Db 713 DFGPPEAALPAAPGLPPTPVSIDIVVPSLSGSGQVDAGLVPAPABAPLPSTVST 772
Qy 549 -----PG- 557
Db 773 LPSPPPPVKKPPAPPLPRTRLLYTYPDGAKYAGSLFESDCMVLNANPGHPPGGL 832
Qy 558 -----HATAASLPXYRA--TLOAG-- 574
Db 833 CHAFYQRFPAFYSTERIMEGLAAYTLTPRPILHAVA--PDRVQONKRLERAAVRET 889
Qy 575 -----INTDVQOLKITL- 586
Db 890 CSRRGTAAVPLDLSGIYQVPSLSFDMERNHRGDELYTEPAAAWFEANKMPQALTI 949
Qy 587 -EHALKT-----ID-----GITLSPVGLMEY-GRPQSGKGTILALAE 626
Db 950 TEDTARALNALIDATEVGRACAGCTISP--GIYVQPTAGPGSGKRSI-----QQ 1002
Qy 627 GGRALVYAPTRERLEAMDRRIKPPSAGATQVALAIL--RRATVAGAPFATVIVDECFMF 684
Db 1003 GDVAVVAVPTRELNSWRRR--GFAFTPTHTAKRVTTGR-----VIDEAPSL 1049
Qy 685 PLVYVAIVHALSPSRIVLVGDVHQLGIFIDFOGTSANMPLVRDVVKOCRRRTENQTRCP 744
Db 1050 P-PHLLHLH-MQRASSVHLLGDPRQIPALDFEHAGL-VPAIRP--ELAPISMVHTHRCF 1104
Qy 745 ADVVATTFPQSLYPGCTTSGCVASISHVAPDRYNSAQTLCTQOEKSHHGAEGANTVH 804
Db 1105 ADV--CELIRGAYPKIQTTSRVRSRFLWNEP--ASGOKLVFOAKAAN--FGAITVH 1156
Qy 805 EAGGRFPASVILHYNSTAEOKLAEKS--HLVGIIRHNLHYIRPCTDIEROLNHS- 861
Db 1157 EAGGATFTETTTI--ATAADARGLIQSSRAHAIVALTTRHTEKCIILDPG-LKEIGISD 1212
Qy 862 -----AKAEVFTDIPADLEITTVKSEVOGRN-EVMATIPPOSATPGAILHLKNEF 912
Db 1213 VIYNNPFLAGGVGHRR-----SVTPRGNPQNLGTLQAFPPSCOI--SAYYQLAEEL 1264
Qy 913 GDOPDCCGVALAKTGVFVFGRAKINVELAPDAPTRPHRAFOEGVQWVKVN----- 965
Db 1265 GHRP-APVAIVLPPELEQGLLYMPQELTVSDS-----VLVFEILYNNVHCMP 1312
Qy 966 -ANNGKQALOTLSRYKRSADLPHE-AKEDVYGRMINSIDRHMDVWTEADARAVFET 1023
Db 1313 APQORAAVSTLVERGRT--KLYEAHSDVR--ESLARFPTIGVQVQATTCCELVEL 1366
Qy 1024 QLFQTOGAGVEDLEPD--DPYIRIDIFLMKTQOK-VSPKPIITGVGGIGIAHKSILN 1080
Db 1367 VEMAVBEKGDSANVLELDLCNRDVSRTTFQKDCNKRTTGRTIANGVGGISMSWTFPC 1426
Qy 1081 FVLAATRIIB-EILRTGSTRVYSNGLPDBEAMLEAKINOVPHATFVADMTERTDTA 1139
Db 1427 ALRGPWRAIEKETLIPPIFYG--DAYEESVPAVAASGAGSCMVEPNFSEPDST 1482
Qy 1140 HNTSELFPAALEIRITPAAAVVLPFRROGKRLRA-----KOLGSEVVDGLDSGAAM 1194

```

```

Db 1483 QNNFSLGBCVWMBEOMPMWMLRLYLHVSAMWILQAPKESLGFPMKKH-----SGEPG 1536
Qy 1195 TPCRNITTFSAAMVTL--FRGVKFAAPKGDG--LLCGSHYLRDASRLHMGERYKTKHLK 1251
Db 1537 TLLMNTIMWNAIITAHCYBEFFRVAAPFKGDDSVLCSDYRQSHMAALTAGCGLK--LK 1593
Qy 1252 VEVOKIVPYIGLIVLSAQVTL--DPVRS 1278
Db 1594 VDYRPIGLYAGVVAAPGLGTLDPVRF 1621

RESULT 8
ID Q8JUN3 PRELIMINARY; PRT; 1707 AA.
AC Q8JUN3;
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE ORF1 protein.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OX NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JDK-sap;
RA Mishiro S.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=JDK-sap;
RA Takahashi K., Kang J.-H., Ohnishi S., Hino K., Mishiro S.;
RT "Genetic Heterogeneity of Hepatitis E Virus Recovered from Japanese Patients with Acute Sporadic Hepatitis.";
RJ Infect. Dis. 185:1342-1345 (2002).
RL EMBL: AB074817; BAB93538.2; -.
DR InterPro: IPR002589; ALPP.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR007095; RNA_Pol_DS_PS.
DR InterPro: IPR007094; RNA_Pol_Polr.
DR InterPro: IPR000606; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransf.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransf; 1.
DR SMART: SM00506; ALPP; 1.
DR PROSITE: PS50507; RDRP_POSITIVE; 1.
DR PROSITE: PS50521; RDRP_VIRAL; 1.
SQ SEQUENCE 1707 AA; 186913 MW; EALAC783BA32C7 CRC64;

Query Match 5.9%; Score 525; DB 12; Length 1707;
Best Local Similarity 21.4%; Pred. No. 8.2e-26;
Matches 377; Conservative 172; Mismatches 578; Indels 638; Gaps 72;

2 YAAATDVAAVAAADVAAVAVNLQAAVKLDPAFLKALFT--LRLYTP--LAPKGGTL 57
6 FIAAPGVTAIEQAAALAAANALANAVV--RPLSRLQTRILINLMQPROLVERPEVL- 62
58 PTOHPIAGHORAEBVLNHPARGS--TVLEIGPSLSALKHGAAPAPADVHG-C-TRY 115
63 -NNHPI---QRYIHNELEQYCARGRCLVEGAHPRS--INDNPV---LHRCFLKP 110
116 GTDGSRHITALE-----SRGVATGRPEPKADASLIANGIASRFFCVGVGSCAFKSR 168
111 VGRDVGRWYTAAPTRGPAANCRRSALGLRP-----ADRTYCFRGFGCTTAAE 158
169 VGIANSLYDVTBELANAPENHGLHVRAPFMGPBEELYMNVVNAELGYRHYVIREPM 228
159 TGIATLSIHMLPADVAAEAMARHGMTRLYAALHPPEVLLPPTGYHTTSYLILH----- 212
229 AVWDCAFGQGDLLHHPPELDFINBSQERRIERLAAAGSYRRAVITPBGDDDKWDAYLHDF 288

```

```

Db 213 -----DGD-----RAVITYEDSSAG--YANDV 233
Qy 289 HTWLAVLVNRYPPFGSLHIEVORRHGSSIELRTRAP-----PGDMLAVNPT 340
Db 224 SILRAM--IRTKVTGDHPLVIERVAVGCHVLLTLTAAPBSPMPVYPPESTEVYRS 291
Qy 341 SOGLCRINIFYNADASGTEHKTILTSQHKVMMLNFMQTRBEKLVDMTVLMSAPARL 400
Db 292 IFPGSGSPSLF--PTACSTKSTFHAVPVHIMDRMLFGATLDDQFC--CSRLMTYLRGIS 348
Qy 401 RAIVASEVTESSMNI SPADLVRTVSLVLIHIERRAVAV-----443
Db 349 YKTVGALVANEGMNASDALTAIVTAAYLTICHQRYLTQALISGMKRLBLEHAKRIT 408
Qy 444 -----KTAKDVG--ETSPWESLKVHLS-----466
Db 409 RLVSWLFKSGRDYIPGRQLQFYACCRWLSAGPHLDPRVLVFEBAAPCRSLRKXAH 468
Qy 467 --CCGLRNL-----KGTDV-----478
Db 469 KFCCEMRWLQDDCTCFLQPVGEVGEQGYDNEAFEGSDVDPAEBAVTSISGSYVTGSL 528
Qy 479 -----VFTK-----482
Db 529 QPLVQALGIPSDLABRGLTATVSVSDTGRLTCKTKTMKNTFTTFTTIDQAQLEANGPE 588
Qy 483 -----RVNDK-----487
Db 589 QVLSFDLAKQTMAGPHSLSYALTAPGLEVAVVNSAGLDCRAVPSPGAVATPSAAGEVSAP 648
Qy 488 --YRV-----HSL-----493
Db 649 CSALYRNRQVGRHSLIGLWYTPBGLIGLPPFAPGHTWESANPFCGSESTLYTETWASVS 708
Qy 494 -----GDIICDVRLSPEDQGLPSRVPPARV--HDSRELEVLR 530
Db 709 GFSSCFSPPEPALDPPAEADTPMAVDVPPATLTLPQAPAFRAVPPQLADGDVAR 768
Qy 531 EA-GCYNERVP--STEPVEE-----548
Db 769 ASPGVSAAPVPAQSVTDPPVGRRLHLHTYPDGSKVYAGSLPESCTMLVNANPGHRG 828
Qy 549 -----PGQD-----ADMATAASLEBYRA-----569
Db 829 GGLCHAFYQRFESFDPAPFVNSDPAAYTLTPRPIHAVA--PDYVENHPKYLEAY 885
Qy 570 -----TLQAGLNTD-----578
Db 886 RETCSRGTAAYPLLAGIKYKVPVGLSPDAMERNRPGDELTYLTERALAMFEANRPTLPA 945
Qy 579 -----VKQLKITLENAL--KTIDGLTSLPYRGLEMYE--GPPSGSGKTGLIAL 623
Db 946 LTITEDTARTANLALBDSATEVGRACACGRVBP--GVVHYQFTGVPESGSRSI-----999
Qy 624 EAAGKALYVAFTRLEIREAMDRIRKPPASATQHYALAILRPAITBGAFFATVVIDECM 683
Db 1000 -OQGEVDVVVPTRELRNSMRBR--GFAAYTPHRAARTRG-----RRVVIDEAPS 1047
Qy 684 PLYVVAIVHALSPSRIVLYGDVHQIGRIDFGSTANMLVADVVKQCRRTPTQTKRC 743
Db 1048 LP-PHLLIH--MQRASSVHLDBPNQIPADIEHAGL-VFAIRP--ELVPTIMMLITHRC 1102
Qy 744 PADVATTFPQSLYPCCTTSGCVASISHVADYRNSAQTLCTFOEKSRRHAGSAMTV 803
Db 1103 PADV--CELIRGAYPKIGTASRVFSLFMGBEPV--GQKLVFQAAKAAAN--PGATIV 1154
Qy 804 HEAGRTPASVILHNGSTAEOKLAEKS--HLVGIPTHTHMLYIRDPGTGIEBQJLMS 861
Db 1155 HBAQATFETETIT--ATAADRGILQSSRAHAIYALTHTKCVVAVDAG--LBRVIGIS 1210
Qy 862 -----AKAEVFTDIPAPLEITTVKSEEVQRN--EVMATIPQSATPFGAHILHRKN 911

```

```

Db 1211 DAIVNPFELSGGQIGCRP-----SVIPROCTINNVDTLDAPP--SCPSAHQJLAE 1262
Qy 912 FGDQPDGCVALATGTYEFGGRKINVELAPDTPRPHAPQ--EGVQWYKVTNASKH 970
Db 1263 LGRP-APIAVLPCECELEQGLYMPQELTSDSV-----LTFELTDIVHCRMAAPSORK 1317
Qy 971 QALOTILSRKTSADLPLHAKS--DYKRMNSLDHRMDVTVTDABRAAFYQJLKTQ 1029
Db 1318 AVLSITLVGRYGRK--KLTYBAHADVGRSLNHFTIPBLGVPVSVTTCELYELVEMAVEGQ 1374
Qy 1030 RGTVEEDLLEPDDPYIRD--IDFLMKTQOK-VBPKPINTKVGQGIANHSKSLNFVLA 1085
Db 1375 DGSANLE-----DLCSDVSNITFPQDCKNTFTGTGTIAHKVQGISANSKTCALFGP 1430
Qy 1086 WIRLE--RIFTGSRTRYNSGLPDDEBAMLEAKINQVPATVVSADWTEPDTHANNTS 1144
Db 1431 WFAIRIKELIYLAIPNVPYG--DAVEDTVLAAVAAGAPCKVENDPFSBFDSTQNNFS 1486
Qy 1145 ELPAALLERIGTPAANVLER-----BECGRITLAKGLSGVEVDGLDGAAMTPC 1197
Db 1487 LGLCITIEBEGMPQMMIRLYHLVRSAMVLDAPPESLR--GPWKCH-----SGEPGTL 1538
Qy 1198 RNTFSAVMULT--FGVKEPAFKGDS--TLQSHYLRFPASRLHMGERYKTGHLKVEV 1254
Db 1539 NMTVMNNAVIAHCYBFDLKVAAFKGDSVVLCSDYQSRDAAALLIAGCGLK--LKVD 1595
Qy 1255 QKIVPYIGLVASBOVVL--DPVRS 1278
Db 1596 RPGLYAGVVAVAPGLGTLPDVVRPA 1620

```

RESULT 9

Q8JUN0 PRELIMINARY; PRF, 1709 AA.

AC Q8JUN0;
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DS ORF protein.
 OS Hepatitis B virus.
 CC Virusess; sRNA positive-strand viruses, no DNA stage;
 CC Hepatitis B-like viruses.
 OX NCBI_Taxid=12461;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JKN-Sap;
 RA Miehito S.;
 RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=JKN-Sap;
 RA MEDLINE=21996248; PubMed=12001054;
 RA Takahashi K., Kang J.-H., Ohnishi S., Hino K., Miehito S.;
 RT "Genetic Heterogeneity of Hepatitis B Virus Recovered from Japanese
 RT Patients with Acute Sporadic Hepatitis.";
 RL J. Infect. Dis. 185:1342-1345(2002).
 DR EMBL; AB074918; BAB93539.2; -.
 DR InterPro; IPR002589; A1pp.
 DR InterPro; IPR001788; RNA_dep_RNAPol2.
 DR InterPro; IPR007095; RNA_pol_DS_P8.
 DR InterPro; IPR007094; RNA_pol_PSVir.
 DR InterPro; IPR006066; Viral_helicase1.
 DR InterPro; IPR002588; V_methyltransf.
 DR Pfam; PF01443; RNA_dep_RNAPol2; 1.
 DR Pfam; PF01660; Vmethyltransf; 1.
 DR SMART; SM00506; A1pp; 1.
 DR PROSITE; PS50507; KDRP_POSITIVE; 1.
 DR PROSITE; PS50521; KDRP_VIRAL; 1.
 SQ SEQUENCE 1709 AA; 187485 MM; 1C9F591068860971 CRC64;

Query Match 5.9%; Score 525; DB 12; Length 1709;
 Best Local Similarity 21.6%; Pred. No. 8.2e-26;

Matches	396;	Conservative	182;	Mismatches	552;	Indels	700;	Gaps	83;
Qy	13	AAADVAVANNVQOBAVKLDFAFPRLKALET--LHRLYYE--LRFKGGTLPTQHPILAQH	Q	68					
Db	22	AAANSALANAVVVR-----PFLSRVQTEIILINMQRLQVFPREVL--WNHP-----Q	D	68					
Qy	69	RVAEEVLINPARGGS--TVLEIGPSLSHSLKLGAPNAVVDYHGC--TYGTRDGSRIHTA	Q	126					
Db	69	RVINNELEQYCRAAAGRCLEVGAMPRS---INDPNV---LHRCFLPVRGRDVGQWYSA	D	121					
Qy	127	LE-----SRSVATGPREFKADSLANGIASRPECVDVGSSCAFKSRVGIANHSILDV	Q	179					
Db	122	PTGRPANCRSALRGRLPP-----ADRTYCFDGSRCFAAETGVALSLHDL	D	169					
Qy	180	TLIEELANAFENHGGLMVAFAFMHMEPELLYMDNVVAELGYEFHYIEEPMAYVDCAFQGD	Q	239					
Db	170	MPAVVAEAMAHGGMTRLVAAHLHPREVLLPQTHHTSYLLH-----DGD	D	215					
Qy	240	LRLHFPPELDPTINESOERRIERLAARGYSRRRAVIFSGDDMDGADLHDFHTWLAVL---	Q	295					
Db	216	-----RAVVTVEGDTSGAG--YNHDSILRAAIRTK	D	244					
Qy	296	LVRVYPTPPGFSLHTEVORRHGSSIELRITRAP-----PGDMLAVNRTSGGLRI	Q	347					
Db	245	IVGHP-----LVIERVAIGCHFVLLITAAPESPMPYVPYPRSTEVYRSIFPGGS	D	298					
Qy	348	PNIFEYVADASGTEHKTILTSQHKVMMLNFMQTRPEKELVMTVLMSFARALRAIVAS	Q	407					
Db	299	PSLF--PACGCTGKTFHVAHPVHIWMRLMFGATLDDQAF-C-CSRMTYLRGISTKYVTYGA	D	355					
Qy	408	EVTSSWNI SPADLVRTVVSLSVLIHIIERRAAVAV-----	Q	443					
Db	356	LVAENGMNASDATALVAVITAAVLTICHQRYLRTQASIKGMRLEVEHAQKFTRLYSWLP	D	415					
Qy	444	-KTAKDDV-----FGESTFMES-----LKHVLGS--CCCLR	Q	471					
Db	416	EKSGRDVTPGROLOFYACCRRLSAGFLHDPRLVDEAVPCRCRTFLKVAKGKCFEWR	D	475					
Qy	472	NL-----KGTDV-----	Q	478					
Db	476	WLQGECTCFLEPAEGLIDGQHDNEAYGSEVDPAEPHLDVSGTYAVHGHQLEALYAL	D	535					
Qy	479	-----VFTKRVVDKXV-----	Q	490					
Db	536	NVPHDIARASRLTATVELVASPDLBCRTVLGNKTFPTTVVDGHLFANGPEEYVLSFD	D	595					
Qy	491	-----HSL-----GDI--ICD-----	Q	499					
Db	596	ASRQSMGASHSLTYELTPAGLQVRISSNGLDCTAVPPGGAPSAAPGEVNAFCAALYRY	D	655					
Qy	500	-----VRLSPQV-----GR-----	Q	509					
Db	656	NRFQORHSLTGGLMLHPGLIGIFPPSPFGHIMEBANPCCGGLTYRTWTSIGFSSDPS	D	715					
Qy	510	-----LPSRVPPARVHFDEBEVL--REAGCYMERPVPSPT--	Q	544					
Db	716	PPEAAAPVPAAPAGLHPHPTPV-----SDLWVLPPESEBSQIDAAPVPVPKVTGLSP	D	769					
Qy	545	-----PVEPQG-----FDADL-W-----	Q	557					
Db	770	IVLAPSPPLPSPVRKPPSPPSRTRRLLYTPDGARVYAGSLFESDCMLVNASNPGHRP	D	829					
Qy	558	-----HATPAASLPEYRA-----TLQAG	Q	574					
Db	830	GCGGLCHAFYQRPPEAFYPTBEFLMEGLAAYTLTPRPITIHAVA--PDYRVBONPRLEBA	D	886					
Qy	575	-----LNTDVQQLKITL-----	Q	586					
Db	887	YRETCSRGTAAVPLLGSGIYQVPVSLSFDAWERNHRPGDELYLTERPAAMPEANKPSP	D	946					
Qy	587	-----ENALKT-----ID-----GLTSPVRLGEMTE--GPPGSGKTGTLAA	Q	622					
Db	947	ALTITEDTARANIALLEIDAAITEVGACAGCTISP--GVNHQPTAGVGGSGKSI--	D	1001					

QY	623	LEAAGKALVYAPRREIRREAMDRIKREPSASQCHV--LALIRRAAAGAPRATVYIDE	680
Db	1002	--QGDVDVAVVPPRREIRRENSMRR--GFAAPRPHAAATYIGR-----VVIDE	1046
QY	681	CFMEPLVVAIVHALSPESRIVLVGDVHVGICFIDPQSTANMPVLRDVHQCRRRTENQ	740
Db	1047	APSLP-PRILLIH-MQRASSVHLLGDNOQPAIDFHAAGL-VRAIR-ELAPISWHAVT	1101
QY	741	KRCPADVATTFQSLYPCGCTTSSGCVASISHVAPDRNSQAQTLCTQDEKRNHGAEGA	800
Db	1102	HRCPADV--CELIGAVAPKIQTSRVRLSIFMNEP---AIGQLVFTQAQAKAN--PGA	1153
QY	801	MTVHEAGGRPFASVYILHNSSTAQKLLAKRS--HLVGTTRTNHLYINDPPGDIEROL	858
Db	1154	ITVHEAGATFETTTT--ATADARLEISSRAHVALVATRTREKCVILLDARG-LIREV	1209
QY	859	NHS-----AKAEVFTDIPARLEITTVKSPSEHYQRN-EVMATIPQSGATPHGAITHL	908
Db	1210	GIDSVYNNPFLAGSGVGHNR-----SVIPGNPNQNIQTLQAPRPSQI--SAVHOL	1261
QY	909	RKNFGDQDQCCVALATGYEVPFGRAKINVELAEPRATKPRHARQ-EGVQWVKVTNAS	967
Db	1262	ABEILGHRP-AVVAVLPRCEBELGGLYMPQELTVSDV---LVFELTDIVHCMAAPS	1316
QY	968	NKHQALOTLISRYKRGADLPJRH-AKEDVGRMNSLDREMDTVTEDADRAVFERQK	1026
Db	1317	QRKAVLSTLVGRYKRT--KLVEAHSQVDR--ESLARPIPTIGVQATTCBELVELVEA	1370
QY	1027	FTQRGTVBDLLEBD--DPYRIDIDPLAKTQOK-VSPKPINTGKVGQIAHKSILNVL	1083
Db	1371	MVEKGQDQSAVLELDLGNQDVSRITFPQDQCNKTTSETLHNGVVGIGIAMSXTPCALF	1430
QY	1084	AAWTRILE-ELTRGSTRVRYSNGLPRBEEMLEAKINOVPHATFVSADWTEPDYAHN	1142
Db	1431	GPWFRALIEKELTALLPPNIFVG---DAYESVPAAAVSGAGSCMVEPNFSEBDSYQN	1486
QY	1143	TSELLPALLERIGTIPAAVVLPRERCGKQTLRA-----KGLSVEVDGLDGAAMTPC	1197
Db	1487	FSLGECVWHECGMPQMLIRLYLVASAMTLOPKESLKGPMKCH-----SGEPETLL	1540
QY	1198	ENTIPSAVMLTL--FRGVKPAAPKGDSD--LTCGSHYLRPADSRILHGERVYTKHLKVEV	1254
Db	1541	MNTVMNMAIIGHCBFRDPFRVAPKGDSDVVLGSDYQNSNMAALINGQGL--LKVDY	1597
QY	1255	QKIVPYIGLVASAEQVVL-DPVSAKIFGRCCTSELLYSKYVEADVDTKGSADARYHS	1313
Db	1598	RPIGLYGVVVVAPGIGTLPDVVRPA---GR-LSE-----KNWGP-----	1632
QY	1314	LLCHMSACYVNVAPESAAVYIIDAVVRGRG	1343
Db	1633	-----GPERABQRLAACVDFLRG	1650


```

QY 1085 AMIFLE-ELIRTSRTYRYSNGLPDEEAMLEAKINQVPHATFVSADMTPEPTAHNT 1143
Db 1430 PMFPAIRKEIILAAALAPVFG----DAVEDTVLAAAVAGACVKFENDPSEFSTONPF 1485
QY 1144 SELFPALBERIGTPAAAVNFRERCGKRTLRAGLSVEVDG-----LDSGAMPSCR 1198
Db 1486 SLGLECTIMECGMPQMIRLY-----HLVRSAMITLOAPKESLRGFMKKSHPETILM 1539
QY 1199 NTIFSAVMTLT--FRGVFAFAFKGDS-LTGSHTLFPDASRLMGERYKTKLKVEVQ 1255
Db 1540 NTVMNMAIACHYERDLKVAAFKGDSDVLCSDVRGRDAVAVLGGGLK---LKVDPR 1596
QY 1256 KIVPYIGLVSAEOVLT-DPVRSA 1278
Db 1597 PIGLYAGVVAAPGLCTLPDVVRA 1620

RESULT 12
ID Q8AZM5 PRELIMINARY; PRT; 1709 AA.
AC Q8AZM5;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
DE Genomic RNA, complete genome, isolate:HE-JA10.
OS Hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis E-like viruses.
OC NCBI_TaxID=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HE-JA10;
RX MEDLINE=22447293; PubMed=12560575;
RA Tokita H., Harada H., Gotanda Y., Takahashi M., Nishitawa T.,
RA Okamoto H.;
RT "Molecular and serological characterization of sporadic acute
RT hepatitis E in a Japanese patient infected with a genotype III
RT J. Gen. Virol. 84:421-427(2003).
RL EMBL; AB089824; BAC44897.1; -.
SQ SEQUENCE 1709 AA; 187464 MW; 64B7ACE711713CDA CRC64;

Query Match 5.9%; Score 520; DB 12; Length 1709;
Best Local Similarity 21.5%; Pred. No. 1.8e-25;
Matches 393; Conservative 187; Mismatches 556; Indels 688; Gaps 84;

QY 13 AAADAVANVLCQRAVXLDFAPLKALET--LHRLYLP--LRFKGGTLPTOHPILAGHQ 68
Db 22 AAASALANAAVVR-----PFLSRVQTEILINLMGRQLVFRREV--WMHPI-----Q 68
QY 69 RVABEVLNFAFGRS-TVLEIGPSLSHSAKLHGAENAPVADYHG-TKXTGTRDGSRHITA 126
Db 69 RVIHNELOYCRABRCLEVGAPHS---INDNPNV---LHRCFLRPVGRDVQRYSA 121
QY 127 LE-----SRSVATGRPEFKADSLANGIASRPFCDVGVSQAFKSRVGIANHSLYDV 179
Db 122 PTRGPANCRSRALRGLRP-----VDRTYCEGFSRCAFAAETGVALLSLDL 169
QY 180 TLEELANFENGLHNVAFMFMPEELLYMNVNAALGYRPHVIEBPMAVDCAFQGGD 239
Db 170 WPAVDAEMARKGRTLYAALHLPREVLLPCTYHTTSLYLN-----DCD 215
QY 240 LRLHPELDIFINESQERIERLARAGSYRRAVIFSGDDWDGDAYLHDFHTLAVL--- 295
Db 216 -----RAVVTYEGDTSAG--YNHDSILRAMIRTK 244
QY 296 LVRYNPTPGSLHIEVGRHSSIEIRTRAP-----PGDMLAVVPTSGQLCRI 347

```

```

Db 245 IVDHP-----LVIEVRAIGCHFULILTAAPSPMPVYFIRYRSIEVYRSIFGCGS 298
QY 348 ENIFYVADAGTEHKITLTSQHKVNLNFMQTRPEKELVDMTYLMSPARARLAIVAS 407
Db 299 PSLE--PSACGKSTFAVVPVPHIMDRMLFGATLDDQAF-C SRLMTYLRGI SYKYTVGA 355
QY 408 EYTESWNIPRADIYRVVSLYVLIHERRAVAV----- 443
Db 356 LVANEGNNSADLTAIVTAAYLTICHORYLRTQALSKMRRLVEHAKETLYLWLF 415
QY 444 -KTADQV-----PGTSFWBS-----LKHYLGS-CCGLR 471
Db 416 EKSARDYTPGRQLFYACCRMLSGAHHLPBRLVLPBESVPCRCRTLYLKVAGAFCCMR 475
QY 472 -----NLKSTDVFTKRVVDKRV- 490
Db 476 WLAGGCTFLPAPAGLIGDQHDNEAYEGSEVDPAAPRALDVGSTVAVYGHQLBALYAL 535
QY 491 ----- 490
Db 536 NVPHDIARASRLTATVELVASPDRLRCRTVYLNKTRTYVVDGAHLBANGPEQYVLSFD 595
QY 491 -----HSL-----GDI--ICD----- 499
Db 596 ASRSGMAGSHSLYTELTPAGLQVRISNGLDCTALPPGGAAPAAQGVAAFCALYRY 655
QY 500 -----VLSPEQV-----GP----- 509
Db 656 NRFTGRSLTGLMLHPEGLLGI PPSPGHIWESANPFCBGITLYTRMTSGSSDPS 715
QY 510 -----LPSRVPAVFP-----HDEBEL------VLRAGCYN----- 536
Db 716 PPEADAPAAAPGLPHTPEVSDIWLPPSSSQIDAAPVPPSKTAGLPSPIVLAAP 775
QY 537 ---BRPV--PSTPVBS-----POG-----FDADL-W----- 557
Db 776 PPLSPVAKPSGSPPSKTRLLTYTPGAKYAGSLPESDDMLVNASNPGHRPGGLCH 835
QY 558 -----HATAASLPEYRA-----TLQAG----- 574
Db 836 AFYGRPEAFYPTFIMEKGLAAYTLFPRPIIHAVA---PDYRVEQNPKRLBAVRETC 892
QY 575 -----INTVKKQLKITL-----E 587
Db 893 RRGTAAYPLGSGYGVVSVLSFPAWERNHRPGDELYTEBPAAMFEANKSOPALITTE 952
QY 588 NAKKT-----ID-----GLTSPVAGLEMYR---GPQSGKTGTLIAALBAAG 628
Db 953 DTAKTANLALETIDATBVGACACTISP--GIYHIOPTAGVPSGSKRSI-----QGD 1005
QY 629 KALYVAPTRLEAMDRIRKPSASATQVVA--LAILARATAGAPPAVVIIDECFMPPL 686
Db 1006 VDVVVVPFRELKNSWR--GPAAFPTHATARTIGR-----VVIDEAPSLP- 1051
QY 667 VYVAIVALSRSKVLVGDVHQIGFIDPOSTANMPLVROVVKQCRRTNOTRCAPAD 746
Db 1052 PHLLHLH-MQRASSVTHLGDNQIPALIDFHAQ-VPIIRP--ELAPTSMMHVTYRCAP 1107
QY 747 VVATTFQSLTSGCTTSGCVASISHVAPVDYRNSOAGTLCTQSEKSHHAGAMTVBA 806
Db 1108 V--CELIRGAYPKIOTTSRVRSLFMPNR---AIGQLVTOAKAAN--PGALTYVEA 1159
QY 807 QGRTFASVILHYNGSTABOKLAARKS--HLVGIYTRHTNLVYINDPTGDIROQLNS--- 861
Db 1160 QGATFTEITII---ATADARGLIOSRAHAIVALTRHBKCVIIDAAP-LIREVGISDVI 1215
QY 862 -----AAAEVPTDIPALEITTYKPSSEVQBN-EVATITPQASTPGALHILKXNGD 914
Db 1216 VNNFFLAGEVGHHNP-----SVIPRGNPDQNGITLQAPPSCOI--SAHQLAELBGLH 1267
QY 915 QPDGCVLAKTGYEVFGRAKINVELAPDAPTRPRAFO-EGVQWVKTNASNKOAL 973

```



```

Db 647 NRFTORHSLTGLMLPEGLLGIFFPPEBGIHIESANPFCBEGTLTYRTWSTGFSDFS 706
Qy 510 -----LPSRVPAR-----VPHDBELEVLREACQNERPSPST----- 543
Db 707 PPEAAPAPMAATPGLPHSTPVSIDWVLPPEBSEFQV-----DAAPPPADPAGLFG 759
Qy 544 -----PVEEP-----OCFADL-W----- 557
Db 760 PVLVLPPEPPVHKSIIPPSRNRLLTYTPGAKVYAGSLTESDODLVNANSGHRRG 819
Qy 558 -----HATAASLPETRA-----TLQAG- 574
Db 820 GGLCHAFYQRPPEAFPTFEMREGLAATLTPRIHAVA---PDYRVEQNPKELEMAV 876
Qy 575 -----LNTDVOKOLKITL----- 586
Db 877 RETCSRRTAAVPLGSGIYQVPSLSFDAMERNRPGDELVTPEBANFEAKPQAPV 936
Qy 587 -----ENAKT-----GLTSPVGLMEYB---GPPSGKGTGLTIAL 623
Db 937 LITTEDTARTALALEIDAATEVGRACACTISP-GIVHYQTAGVPSGSKRSI----- 990
Qy 624 EAAGKALVAPTREREMDRRIKPPSASATQOVA--LAILRATAGAPATVVIDEC 681
Db 991 -OQGDVDVVVPTRELNSMRR--GFAAFPTHTAARTIGRR-----VIDEA 1036
Qy 682 FMFPLVVAIVHALSSSRIVLVGDVHOIGFIDFOSTANMPLVNDVVKCRRRTFNOK 741
Db 1037 PSLP-PHLLLLH-MORASSVHLIGDNOIPALDFEHAQ-VPAIRP--ELAPTSMXVTH 1091
Qy 742 RCPADVATTFQSLTSGCAVSIHVADVRNSOAGTLCCTOEEKSRHAGAM 801
Db 1092 RCPADV--CELIRGAPKIOITSRVARSLEFMNER---AIGQLVYTOAKAAN--FGAI 1143
Qy 802 TTHEAGRTFASVITHYNGSTAEQKLAES--HLIVGTRTHNLVIRDPGDIRQIN 859
Db 1144 TTHEAGATFETETIT--ATADARGLIOSSRAHAIVALTRHEKCVIIDAEP-LUREV 1199
Qy 860 HS-----AKAEFTDIPALELTTVYKPESEVQRN-EVMAITPQASATPHGATHLR 909
Db 1200 ISDVIVNPFLLAGEVGHHR-----SVIPRGNPDONIGTLOAPPPSCOI--SAHQIA 1251
Qy 910 KNFGDOPDCCVALATKYEVFEGGRAKINVELAPDATEKPHRAFO-EGVQWYKNTASN 968
Db 1252 EELGHRP-APVAANLPPCELEBGLLMPOELTVSSV---LVFELTDIVHGRMAAPQ 1306
Qy 969 KQOALQTLISRYTKRSADLPLHE-AKEDVKMLNSIDRHMDWTVEDARDRAVETQLKF 1027
Db 1307 RKAVISTLVGRYGRRT--KLVEAHSQVYR--ESLARFIPITGVQATTCELVELEAM 1360
Qy 1028 TORGGTVEDELEPD--DPYIRIDFLMKTQOK-VSKRPINTGVGGGIAHSLNFVLA 1084
Db 1361 VERGGDQSAVLELDLCNRDVSRTTFQCKCNKFTTETIAHGVGGGISMSTFCALRG 1420
Qy 1085 AMIRIDE-EILRTGSRTRYSNGLPDEEAMLEAKINQVPHATFVSADTEBDTANHT 1143
Db 1421 PMRALEKEILLAPNIFIG---DAYESVPAANVSGAGSCMVRENDPSEBDSQNNF 1476
Qy 1144 SELLEALLERIGTPAAAVNLFRERCQKRTLR-----KGLSVEVDGLLDSGAATPCR 1198
Db 1477 SLCLECEVMEBCMPQMLIRLVHLSAWILQAPKESLKGFMKH-----SGEPGLLW 1530
Qy 1199 NTFPSAAMLT--FGVKAFAFKGDS--LLCGSHYLRPDASLHNGERYKTHGLVYEV 1255
Db 1531 NTVMNAAILAHCEYEFDFRVAARFGDSDVVLCSDYQSRNAALILAGCGLK--LVNDYR 1587
Qy 1256 KIAPYIGLVSAREQVTL-DEVRSLAKIFGRCYTSELISKYCAVAVDITKMGSDARVHSL 1314
Db 1588 PIGLVAGVVVAPGLGTLPDVVRFA---GR--LSE-----KNMGP----- 1621
Qy 1315 LCHMSACYNVAPEASAAYIIDAIVVRFGG 1343
Db 1622 -----GPERAEQLRLAVCDPLRG 1639

```

```

RESULT 14
Q9YLR1 PRELIMINARY; PRT: 1708 AA.
AC Q9YLR1
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Polypeptide.
OS Hepatitis B virus.
OC Virusess; ssRNA positive-strand viruses, no DNA stage;
OC Hepatitis B-like viruses.
OX NCBI_Taxid=12461;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US2;
RX MEDLINE=98178637; PubMed=9519822;
RA Schläuder G.G., Dawson G.J., Erker J.C., Kwo P.Y., Knigge M.F.,
RA Smalley D.L., Rosenblatt J.B., Deval S.M., Mushawar I.K.;
RT "The sequence and phylogenetic analysis of a novel hepatitis B virus
RT isolated from a patient with acute hepatitis reported in the United
RT States.";
RL J. Gen. Virol. 79:447-456(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=HEV-US2;
RX MEDLINE=99190429; PubMed=10092008;
RA Erker J.C., Deval S.M., Schläuder G.G., Dawson G.J., Mushawar I.K.;
RT "A hepatitis B virus variant from the United States: molecular
RT characterization and transmission in cynomolgus macaques.";
RL J. Gen. Virol. 80:681-690(1999).
DR EMBL: AF060669; AAD15815.1; --
DR MEROPS: C41.001; --
DR InterPro: IPR002589; A1PP.
DR InterPro: IPR001788; RNA_dep_RNAPol2.
DR InterPro: IPR007095; RNA_pol_D8_P8.
DR InterPro: IPR007094; RNA_pol_P8vir.
DR InterPro: IPR006066; Viral_helicase1.
DR InterPro: IPR002588; V_methyltransfer.
DR Pfam: PF01661; A1PP; 1.
DR Pfam: PF00978; RNA_dep_RNAPol2; 1.
DR Pfam: PF01443; Viral_helicase1; 1.
DR Pfam: PF01660; Vmethyltransfer; 1.
DR SMART: SM00506; A1PP; 1.
DR PROSITE: PS50507; RDRP_POSITIVE; 1.
DR PROSITE: PS50521; RDRP_VIRAL; 1.
DR SQUENCE 1708 AA; 187261 MW; BR0D5EBBBDAB547 CRC64;

Query Match 5.8%; Score 512.5; DB 12; Length 1708;
Best Local Similarity 21.6%; Pred. No. 5.7e-25;
Matches 333; Conservative 186; Mismatches 557; Indels 687; Gaps 83;

Qy 13 AADVAAYANVLOQRAVKDLPAPLKALET--LHRLVYR--LRFKGGTLPPCHPIAGHQ 68
Db 22 AAANSAANAAYVYR-----PFLSRVQYRILINMQPQLVFRPEVL--VNHPI-----Q 68
Qy 69 RVAAEVLAHPAAGRS-TVLEIGPSLSALKAGAPAAVADYHGC-TKYGTDDGSHHTA 126
Db 69 RVTHNELBOYCRARFRCLEVAHPRS--INDNPV--LHRCPLRPGGDVQRMWGA 121
Qy 127 LR-----SRVATGRPEFKADASLLANGIARFPCVVDGVSAPKSRVGTANHSLYDV 179
Db 122 PTRGPAANCRKSAKGLP-----VDRTTCFQPSRCAPAAETGVALYSLHD 169
Qy 180 TLEELANAEENHGLHNVRAFMEMPEBELLYMDNVVNAELGYRPHVIEBPNAVDCAFQGD 239
Db 170 WPAADVAEAAVABRGMTLVAAHLPPVLLPGR-----YHTTSY----- 208
Qy 240 LRHHPLELPINESQRRIRLAAKGSYRRRAVIBSGDDDKGDADYLDHPTWLAIV----- 295
Db 209 LLIH-----DGNRAVATYEGDTSAG--YHHDVSIILAMWIRTTK 244

```

QY 296 LVBNYPTFGFSLHIEVORHSSIBLRTRAP-----PGDRMLAVVPRTSQGCRI 347
 Db 245 IVDGHP-----LVIEVRAIGCFVULLTRAPEPSMPYVPIPRSTEVYVRSIFGGGS 298
 QY 348 PNI FYVADAGTEHKITLTLOHKVMMLNMQTRPEKELVDMTVLNSPABARLAIYVAS 407
 Db 299 PSLEF--PSACSTSTFPAVPHIMDKLMFGATLXDQAF-C SRLMTYMLGISVKTVGA 355
 QY 408 EYTESWNISPADLVRTVSLVYLHITERRAAVAV----- 443
 Db 356 LVANEGNASEDLTAVITTAAYLTIHQRYLRTQALSKGMRLEVEHAOKETITLYSMF 415
 QY 444 -KTAQXDV-----FGETSPWMS-----LKKVLS-CGGLR 471
 Db 416 EKSGRDTIPGRQLOFTAQCRMLSGAPHLXPRXIVPDESVPICRRTLTAKVAGKFCFMR 475
 QY 472 NL-----KGTDV----- 478
 Db 476 WLGQECTCFLEPAGLVGDQHDNEAVEGEVDPABPAHLDVSGTAVHGHQLEALYAL 535
 QY 479 -----VTKRVVDKRV----- 490
 Db 536 NVPHDIAARASRLTATVELVASPDRLRCRTVLGKRTTVDGAILKANGPBEVYLSFD 595
 QY 491 -----HSL-----GDI--ICD----- 499
 Db 596 ASRQSMGAGSHSLTYBELTPAGLOVKISSNGLDCTATPPGQASPAEVAFAFCSALYR 655
 QY 500 -----VLSPEQV-----GFLPSRV----- 514
 Db 656 NRFTQHRSLTGLWLMHEGLLGIFFPPSPGHIMESANPCGEGTLYRTWSTSGSSDFS 715
 QY 515 -----PPRVVFH-DREBLEVLR-AGCYNF--RPV 540
 Db 716 PPEAAAPASAAAGLPYTPPVSDIWLPPSESHVDAASVSPVAPAGLTSPVLTTP 775
 QY 541 PSTPVVEP----- 549
 Db 776 PPPPVKAPATSPPTRTLLYTPDGAKYAGSLXSDCDMLVNASNPGHRPGGLCHA 835
 QY 550 -----QGFDA-----DLMHATAASLPEYRA-----TLQAG----- 574
 Db 836 FYQRPFAFYSTERIMEGLAAYLTRPIIHANA--PDRYVQNKRLBAARETCR 892
 QY 575 -----INTVYKOLKITL-----EN 588
 Db 893 RGTAAVPLGSGIYQVVSLSFDAMERNHRPGDELYTERPAAWFAANKPAQPALTTED 952
 QY 589 ALKT-----ID-----GLTSLPVGLMEYF--GPPGSKTGTILALALAGCK 629
 Db 953 TARTANLALBIDAATEVGRACAGCTISP--GIYHYQFTAGVPSGKRSI-----QQGDV 1005
 QY 630 ALVYATPRELEAMDRIRIKPPSASATQVA--LAILRRATAEGAPFATVVIDECFMPPLV 687
 Db 1006 DVVVVPRFELRNSMRR--GFAAFTHTAARVITGR-----VIDEAPSLP-P 1051
 QY 688 YVAIVHALSPSSRIVLVDVHOGIFIDFGTSANMPLVRDVVYKQCRRTTNQTKRCADV 747
 Db 1052 HLLILH-MQRASSVHLIDGDPNQIPALIDFEHGL-VPAIRP--ELAPTSMMHVTTRCADV 1107
 QY 748 VATTFFOSLYPGCTTSGCVASISHVADVRNSQAQTLCTPQEKSHHGAEGMAYVHAQ 807
 Db 1108 --CELRIGAYPKIQTISRVLNRLWNEP--ALGQKLVFTQAKAAN--PGATVYHAQ 1159
 QY 808 GRTFASVILHNGSTAEQKLAERS--HLVGITRTHNLHYIRPTGDIEROLNHS-- 861
 Db 1160 GATFETTTI--ATAARGLIOSSRAHAIALTRHKECVILDAPG-LAREVISDVIV 1215
 QY 862 ----AKAVPTDIPADLEITTVKPSSEVORN-EVMAITIPQSATPGAIHLKKNFGQ 915
 Db 1216 NNFLAGGEVGHNP-----SVIPRGNPQNLGTLQAFPPSCQI--SAYQQLAEELGHR 1267
 QY 916 PDGCVALLATGVEVFGRAKINVELABRATPRKHAFO-EGVQWVKVTNASKHQALQ 974

Db 1268 P-APVAANVLPCELEQGLLYMPQELTVDSV-----LVBELTDIYHCRMAAPGQRAVLS 1322
 QY 975 TLISRTKSGADLPLEH-AKEDYKRMNLSDRHMWTVEDADRAVFTQLTQFGGT 1033
 Db 1323 TLVGRYGRRT--KLYEAHSDVR--ESLARPIPTGPVARTTCELYELVEAMVEKGD 1376
 QY 1034 VEDLEBPD--DPIRIDPLMTQOK-VSPKPIINTKQGGIAHAKSLNPLVLAAMTRIL 1090
 Db 1377 GSAVLELDLCNRVSNRITTFQKCNKFTGTETIAHKVGGISAMKTFCALGMPFRAI 1436
 QY 1091 E-SILRTGSRVYNSGLPDPEERAMLEAKINOVPHATFVSADMTFEDYHNTSBLFA 1149
 Db 1437 EKETILALPNIYFG--DAYBSVPAANVSGAGSCMPENDPSFEDSQNNPSIGLEC 1492
 QY 1150 ALLERIGTPAANVLFRERCGKTLRA-----KGLSVEVDGLDLSGAANTPCRNTPSA 1204
 Db 1493 VVMEBCGMPQWLIRLHYLVASAWILOAPKESLKGFWKCH-----SGEPOTLLMNTVMNM 1546
 QY 1205 ANVLTIL--PRGVFPAAPKDDG--LIGSHYLRPDASRLHNGERYTKHLAKVQKVPYI 1261
 Db 1547 AITAHCEYFDFRVAAPKDDSVVLCSDYRQXRNALALAGGLK--LKVDRPIGLYA 1603
 QY 1262 GLVSAEQVVL-DPVSAALKIFGRCTSELYSKYEAVADITKGSMDARYHSLLCMSA 1320
 Db 1604 GVVVABGLGTLPPVAFPA-----GR--LSF-----KMGF----- 1631
 QY 1321 CYNVVAPESAAVYITDAVVRGRC 1343
 Db 1632 -----GPERABQRLAVCPDLRG 1649

RESULT 15
 ID QYK10 PRELIMINARY; PRT; 1708 AA.
 AC QYK10;
 DT 01-MAY-1999 (TEMBLrel. 10, Created)
 DT 01-MAY-1999 (TEMBLrel. 10, Last sequence update)
 DT 01-MAR-2003 (TEMBLrel. 23, Last annotation update)
 DE Nonstructural polyprotein.
 OS Swine hepatitis E virus.
 OC Viruses; ssRNA positive-strand viruses, no DNA stage;
 OC Hepatitis E-like viruses.
 NC NCBI_TaxID=63421;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Meng;
 RX MEDLINE=97420774; PubMed=9275216;
 RA Meng X.J., Purcell R.H., Halbur P.G., Lehman J.R., Webb D.M.,
 RA Tzareva T.S., Haynes J.S., Thacker B.U., Emerson S.U.;
 RT "A novel virus in swine is closely related to the human hepatitis E
 RT virus.";
 RL Proc. Natl. Acad. Sci. U.S.A. 94:9860-9865 (1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Meng;
 RX MEDLINE=99030877; PubMed=9811705;
 RA Meng X.J., Halbur P.G., Shapiro M.S., Govindarajan S., Bruna J.D.,
 RA Mahapatra I.K., Purcell R.H., Emerson S.U.;
 RT "Genetic and experimental evidence for cross-species infection by
 RT swine hepatitis E virus.";
 RL J. Virol. 72:9714-9721 (1998).
 DR EMBL; AF082843; AAC97208.1; --
 DR GenBank; AF082843; AAC97208.1; --
 DR InterPro; IPR002589; A1PP.
 DR InterPro; IPR001788; RNA_dep_RNApol2.
 DR InterPro; IPR007095; RNA_pol_DS_PS.
 DR InterPro; IPR007094; RNA_pol_PStir.
 DR InterPro; IPR000606; Viral_helicase.
 DR Pfam; PF01661; A1PP; 1.
 DR Pfam; PF00978; RNA_dep_RNApol2; 1.
 DR Pfam; PF01443; Viral_helicase1; 1.

DR Pfam: PF01660; UmeHyltrasef; 1.
 DR SMART: SM00506; AldP; 1.
 DR PROSITE: PS50507; RDRP; 1.
 DR PROSITE: PS50521; RDRP_VIRAL; 1.
 KM Polypeptidein.
 FT CHAIN 55 237 METHYLTRANSFERASE.
 FT CHAIN 975 1219 HELICASE.
 FT CHAIN 1232 1708 RNA-DIRECTED RNA POLYMERASE.
 SQ SEQUENCE 1708 AA; 187407 MW; 7444552DCD616130 CRC64;

Query Match 5.8%; Score 510.5; DB 12; Length 1708;
 Beet Local Similarity 21.3%; Pred. No. 7.8e-25;
 Matches 389; Conservative 183; Mismatches 558; Indels 699; Gaps 81;

QY 13 AADVAAYANVLOQRAVKDFAAPPLKALET--LHRLYLP--LRFKGTLPPTGHPPLAGHQ 68
 DB 22 AAANALANAAVVR-----PFLSRVQTEILINIMQPROLVPRBEVL--NMHPF-----Q 68
 QY 69 RAASEVLNHFARGRS--TVLEIGPSLHSAKLHGAPNAPVADYHG--TKYGRDGSRHITA 126
 DB 69 RAHNELEQYCARAGCCLEVG--AHPRF--INDNRRV--LHRCPLRVGHDVQRWISA 121
 QY 127 LE-----SRSVATGREPEFKADASILANGIASRTFCVDGSCAFKSRVGIANSLYDV 179
 DB 122 PTRGPAANCRRSALGLRP-----VDRTYCFDGFSCAPAAETGVALYSLHDL 169
 QY 180 TLEELANFENHGLHMTAFMMPBELLYMDNVNNAELGYRPHVIEBPAAVDCAFQGG 239
 DB 170 WPADVAEMARHGMRLYALHILPEVLLPCTYHTSYLIH-----DGD 215
 QY 240 LRHPELDFINESQERIERLARAGSYERAVIFSGDDMDGDAYLHDFTWLAVL----- 295
 DB 216 -----RAVYTEGDSAG--YHNDVISILRAMIRTK 244
 QY 236 LVNRYPTPEGSLHIEVORRHGSSIELRTTRAP-----PGRMLAVVPTSGQLCRI 347
 DB 245 IYGDHP-----LVIERVAIGCHFILLTAAPBPSMPYVPRSTEVYVNSIFQPGGS 298
 QY 348 PNIFFYADASGTEHKTILTQKVMNLMNFMQTRPEKELVDMTVLMSFPAALRAIVAS 407
 DB 299 PSLF--PSACSTKSTFFHAVPHVIMWMLFGATLDDQAF--CSRLMTYLRGISYKTVGA 355
 QY 408 EYTESWNISPADLVRTVSVLYVLIIEERRAAVAN----- 443
 DB 356 LVANEGMNSBDALTRAVITPAALYTICHQYLRTOAISKGRLEVEHAKPTIRLYSWLP 415
 QY 444 -XTAKDGV-----FGETSFWES-----LKHVLGS--CCGLR 471
 DB 416 EKSGRDYIPRQLQFYAQCRRMLSGFHLDPRLVDESVPCKRCRTFLKKVAGKFCFMR 475
 QY 472 -----NLKGDVVTTKAVNDKRV- 490
 DB 476 WLGOECTCLFBAEGLVGDYGHNDNEAVESBVDPAEPAHLDVSGTVAHVGRQLKALRYAL 535
 QY 491 ----- 490
 DB 536 NVPHDIAPASRLTATVELTASPDLECTVLANKTFTTTVDGAHLBANGPQVLYSLD 595
 QY 491 -----HSL-----GDI--ICD----- 499
 DB 596 ASRQSGAGSHSLTYELTLAGLOVRISNGLOCTATPPGGAASAPGVAAPFCSLYRY 655
 QY 500 -----VRLSPQV-----GF----- 509
 DB 656 NRTQRHSLTGLMLPEGLLGIFPPSPGHIWEPANPFCGBGLTYTRTWSGFSDFS 715
 QY 510 -----LPSRVPARVPHDRELEVLARAGCYNERPVSTPVEPQ----- 551
 DB 716 PBEAAPVLAAPAGLPHPTPV-----SDIWLPPPSKESQVDAASVPAPBPAGLPS 769
 QY 552 -----PDADL--W----- 557
 DB 770 IVTLPPPLPVKRPPTPPSRTRRLLYTPDGAKYVAGSLPESDCNMLVNASNPGHRG 829

QY 558 -----HATAASLPEYRA-----TLQNG- 574
 DB 830 GGLCHAFYQRPBPAPYPPFPFIMRBLAAYTLTPPIIHAVA--PDYRVEONPKYLEAAV 886
 QY 575 -----LNTDVKQLKITL----- 586
 DB 887 RETSGRGTAPYPLGSGIYVVPVSLGFDAMERNHRPDDELTLRPAAMFAANKAPQA 946
 QY 587 -----ENALKT-----ID-----GLTSPVKGLEMYE--GPPSGRTGTLIAL 623
 DB 947 LTTEEDTARTANLALKEIDAATDVGRACGCTISP--GIVHQTPAGVGGSKSRSI----- 1000
 QY 624 BAAGKALYVAPTELRERAMDRRIKPSASATQVA--LALIRBATAGAPATYVDEEC 681
 DB 1001 -QGGVDVVVPTBELKNSMRER--GPAATFHTAARVTIGRR-----VVIDEA 1046
 QY 682 FMFPLVVAIVHALSPSSRIYLVGDVHQIGFIDFQTSANMPYADVVKOCRRRTFNQTK 741
 DB 1047 PELP--PHLLH--MQBASVHLDDPNQIPALDFEHALG--VPAIRP--ELAPTSMMVHTH 1101
 QY 742 RCPADVATTPFQSLYRQCTTSGCVASISHVADYRNSQOTLCTQDEBSRHGASGM 801
 DB 1102 RCPADV--CELRIGAYPRICQTSRVLRSLEFNNRP--AICQKVFTQAAKAA--PGAI 1153
 QY 802 TVHBAQRTFASVILHNGSABOKLAERS--HLVGTIRTHLYLRDPTGDIERQDN 859
 DB 1154 TVHBAQRTFETETTTI--ATDARGLIQSSRAHALVALYHTEKCVILDAFG--LKRKVG 1209
 QY 860 HS-----AKAEVFTDIPAPLEITTVKSEEVQRN--EVMATIPQSATPHGAILHR 909
 DB 1210 ISDVIYNNFFLAGSEVGNHR-----SVIRGNRPDNLGLQAPFPECQ--SAYHQDA 1261
 QY 910 KNFGDQPPCCVALAKTGYEVFGBAKINVELADATPKHRAFO--BGVMVKTNASN 968
 DB 1262 BELGRP--APVAAVLPCCBELQGLYMPQELTVSDSV--LVFELTDIYHCHMAAPSQ 1316
 QY 969 KQALQTLISRYTSRDLPLHE--AKEDYKMLNSLDHMDMTTEDARDAVETQKLF 1027
 DB 1317 KNAVLSITVGRIGRKT--KLYEAMHSDVR--BSLAFPIITIGPVQTTGELVELVEM 1370
 QY 1028 TORGGTEVDLEPD--DPIYRIDIDFLMKTQOK--VSPKINTQKVGQGIASKSINPYLA 1084
 DB 1371 VBKQDGSAAVLEBLCNRDVGRIITFPQDKCKFTTGERTIANGKQGGISAMSKFCALFG 1430
 QY 1085 AMIRILE--EILRTGSRTRYNSNGLEPDEEAMLEAKINQVPHATFVSADMTETPAHNT 1143
 DB 1431 FMFRAIERKEIILALLPPIFYG--DAYESVFPAAVAGAGSCMVFENDPSBFDSTQNNF 1486
 QY 1144 SELLEPALLERIGPAAAVNLPRRCGRKTLRA-----KGLGSVEVDLIDLSGAWTPCR 1198
 DB 1487 SIGLECVMBECGMQMLIRLYHLVRSAMILLQAPRESLNGKPKKH-----SGEBGTLIM 1540
 QY 1199 NTIFSAAYMLTL--FRGVKPAAPFKGDDS--LTCGSHYLFPAASRLMGERYKTKHLKEVQ 1255
 DB 1541 NTVMMAIIIAHCYERDRVAAAFKGDSDSVLCSVDROGRMAAALLAGGK--LKVDYR 1597
 QY 1256 KIVPYIGLLVASAQVVL--DPVRSALKITFGKCTSELLYSKYVEAVRDITKMSDARIHSL 1314
 DB 1598 PIGLYAGVVAAGLCTLPDVVFA--GR--LSE-----KMWGP----- 1631
 QY 1315 LCHMSACTYNNAPESAAYIIDAIVPFRG 1343
 DB 1632 -----GPERAEQLRLAVCDPLRG 1649

Search completed: January 15, 2004, 16:51:59
 Job time : 95.131 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 16:47:33 ; Search time 28.9919 Seconds
(without alignments)
2486.821 Million cell updates/sec

Title: US-09-991-262-40

Perfect score: 8832
Sequence: 1 MYAKATDVAAVVAADVAA.....LKRIRSDSPVVKISKA 1704

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

1: Issued Patents AA:*
2: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/6CTUS.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8832	100.0	1704	US-08-485-355B-40	Sequence 40, Appl
2	1444.5	16.4	1233	US-09-194-613-5	Sequence 5, Appl
3	510.5	5.8	1708	US-09-462-606-2	Sequence 2, Appl
4	500	5.7	1693	US-08-840-316-1	Sequence 1, Appl
5	500	5.7	1693	US-08-809-523-1	Sequence 1, Appl
6	500	5.7	1693	US-08-471-971-1	Sequence 1, Appl
7	500	5.7	1693	US-09-402-776-1	Sequence 1, Appl
8	500	5.7	1693	PCT-US93-08849A-1	Sequence 1, Appl
9	500	5.7	1693	PCT-US93-08849-1	Sequence 1, Appl
10	479	5.4	1693	US-08-478-507-7	Sequence 7, Appl
11	479	5.4	1693	US-09-128-275A-7	Sequence 7, Appl
12	479	5.4	1693	US-09-553-427-7	Sequence 7, Appl
13	250.5	2.8	2205	US-08-093-453B-2	Sequence 2, Appl
14	213	2.4	1390	US-08-770-544-2	Sequence 2, Appl
15	213	2.4	1390	US-09-579-259-2	Sequence 2, Appl
16	198	2.2	431	US-08-478-507-2	Sequence 2, Appl
17	198	2.2	431	US-09-128-275A-2	Sequence 2, Appl
18	198	2.2	431	US-09-553-427-2	Sequence 2, Appl
19	174.5	2.0	2639	US-09-080-963-3	Sequence 3, Appl
20	168	1.9	2512	US-08-801-263A-9	Sequence 9, Appl
21	168	1.9	2512	US-09-102-248-9	Sequence 9, Appl
22	168	1.9	2512	US-09-367-764-9	Sequence 9, Appl
23	163.5	1.9	1060	US-08-911-393-2	Sequence 9, Appl
24	153.5	1.7	2137	US-09-134-001C-4463	Sequence 4463, Ap
25	152.5	1.7	254	US-09-252-991A-31621	Sequence 31621, A
26	150	1.7	322	US-09-252-991A-31608	Sequence 31608, A
27	148	1.7	2628	US-09-413-814-11	Sequence 11, Appl

28	147.5	1.7	2500	US-08-801-263A-2	Sequence 2, Appl
29	147.5	1.7	2500	US-09-102-248-2	Sequence 2, Appl
30	147.5	1.7	2500	US-09-367-764-2	Sequence 2, Appl
31	146.5	1.7	2517	US-08-801-263A-5	Sequence 5, Appl
32	146.5	1.7	2517	US-09-102-248-5	Sequence 5, Appl
33	146.5	1.7	2517	US-09-367-764-5	Sequence 5, Appl
34	145.5	1.6	1125	US-09-513-783A-152	Sequence 152, App
35	145.5	1.6	1610	US-09-513-783A-22	Sequence 22, Appl
36	145	1.6	907	US-08-783-774-2	Sequence 2, Appl
37	145	1.6	907	US-09-328-599A-1	Sequence 1, Appl
38	145	1.6	907	PCT-US95-04611A-19	Sequence 19, Appl
39	143.5	1.6	618	US-09-252-991A-27666	Sequence 27666, A
40	143.5	1.6	750	US-09-165-239A-4	Sequence 4, Appl
41	141.5	1.6	2431	US-07-920-281C-2	Sequence 2, Appl
42	141.5	1.6	2431	US-08-466-277-2	Sequence 2, Appl
43	141	1.6	669	US-09-107-532A-6532	Sequence 6532, Ap
44	140	1.6	2004	US-08-375-709-15	Sequence 15, Appl
45	140	1.6	2004	US-08-752-929-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-485-355B-40
Sequence 40, Application US/08485355B
Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Christlan, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Flehr Hobbach Test Albitron & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacatlin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RT/DS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 1704 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 40:
US-08-485-355B-40
Query Match 100.0%; Score 8832; DB 3; Length 1704;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 1704; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYAKATDVARVVAADVAAYANVLQORAVLIDPAPLKALETLLRLYYPLRFKGTLLPPTQ 60
DB 1 MYAKATDVARVVAADVAAYANVLQORAVLIDPAPLKALETLLRLYYPLRFKGTLLPPTQ 60
QY 61 HPIILAGHORVAEEVLTNFAFGRSYVLEIGPSSLHSLKLGAPAPADYHGTCTKGTGTRDG 120
DB 61 HPIILAGHORVAEEVLTNFAFGRSYVLEIGPSSLHSLKLGAPAPADYHGTCTKGTGTRDG 120
QY 121 SRHITALSRSVATGPRPEKADASILANGIASRTFCVDGSCAFPSRGIANHSLYDVT 180
DB 121 SRHITALSRSVATGPRPEKADASILANGIASRTFCVDGSCAFPSRGIANHSLYDVT 180
QY 181 LELANAFENHGLHWRAPFMHMBELLVMDNVNNAELGYRPHVIEBPMAVKDCAFQSGDL 240
DB 181 LELANAFENHGLHWRAPFMHMBELLVMDNVNNAELGYRPHVIEBPMAVKDCAFQSGDL 240
QY 241 RLHAFELDFINSOERRIERLAAAGSYSRRAVIFSGDDDMGDAYLHDFHTMLAYLLVRY 300
DB 241 RLHAFELDFINSOERRIERLAAAGSYSRRAVIFSGDDDMGDAYLHDFHTMLAYLLVRY 300
QY 301 PTFPGSLHIEVORRHGSSIELRTTRAPGDMALAVPRTSQGLCRIPNIFYAADSGTE 360
DB 301 PTFPGSLHIEVORRHGSSIELRTTRAPGDMALAVPRTSQGLCRIPNIFYAADSGTE 360
QY 361 HKTILTSQHKVMLNFMQTRPEKELVDMTVLMSFARALRAIVASEVTESSWNISPAD 420
DB 361 HKTILTSQHKVMLNFMQTRPEKELVDMTVLMSFARALRAIVASEVTESSWNISPAD 420
QY 421 LVRTVVSGLYVLIIEERRAAVAKTAKDVFEGTSSWESLKVYLSGCCGLRLKGTDFVF 480
DB 421 LVRTVVSGLYVLIIEERRAAVAKTAKDVFEGTSSWESLKVYLSGCCGLRLKGTDFVF 480
QY 481 TRRVVDKRVVHSLGDIICDVRLSPEQVGLPSRVPARVPHDEBELVREAGCNERYV 540
DB 481 TRRVVDKRVVHSLGDIICDVRLSPEQVGLPSRVPARVPHDEBELVREAGCNERYV 540
QY 541 PSTPVEEBPOGFADLMHATAASLPYRAATLQAGLNTDVOKLTLENAKLTIDGLTSP 600
DB 541 PSTPVEEBPOGFADLMHATAASLPYRAATLQAGLNTDVOKLTLENAKLTIDGLTSP 600
QY 601 VEGLEMYEGPSSGKGTGTLIAALEAAGKALYVAPTRERLEAMDRRIKPPSASATQVAL 660
DB 601 VEGLEMYEGPSSGKGTGTLIAALEAAGKALYVAPTRERLEAMDRRIKPPSASATQVAL 660
QY 661 ALIRBATAGAPATVVIDECMPPLVYVAIVHSLSSSRIVLVGVHOGIFIDFGTSA 720
DB 661 ALIRBATAGAPATVVIDECMPPLVYVAIVHSLSSSRIVLVGVHOGIFIDFGTSA 720
QY 721 NNPLVADVVKOCRRRTFNQTRCPADVATTFEPOSILYPCCTTSGCVASISHVAPDYRNS 780
DB 721 NNPLVADVVKOCRRRTFNQTRCPADVATTFEPOSILYPCCTTSGCVASISHVAPDYRNS 780
QY 781 QAQOTLCFTOEKSRHAGBAMTYHEAQGTTPASVILHNGSTABOGLAEKSHLVGTR 840
DB 781 QAQOTLCFTOEKSRHAGBAMTYHEAQGTTPASVILHNGSTABOGLAEKSHLVGTR 840
QY 841 HNNHLYTRPTGDIIEQLHNSAKAEVFTDIPAPLETITVPSSEVORNEVMAITIPQSAT 900
DB 841 HNNHLYTRPTGDIIEQLHNSAKAEVFTDIPAPLETITVPSSEVORNEVMAITIPQSAT 900
QY 901 PHGAIHLILKKNFGDQDCCVALAKTGYEFGGRAKINVELAPDAPTPKHAFQGVOM 960
DB 901 PHGAIHLILKKNFGDQDCCVALAKTGYEFGGRAKINVELAPDAPTPKHAFQGVOM 960
QY 961 VKTNAASNGAOLQTLISRTYKSAOLPLHEAKEDYKRMINSIDRWMDTJTEDADRAV 1020
DB 961 VKTNAASNGAOLQTLISRTYKSAOLPLHEAKEDYKRMINSIDRWMDTJTEDADRAV 1020
QY 1021 FETQKFTQGGVDELPEDDPYIRDIDFLMKTQOKVSPKPIINTKVGCGIAAHSKSLN 1080

DB 1021 FETQKFTQGGVDELPEDDPYIRDIDFLMKTQOKVSPKPIINTKVGCGIAAHSKSLN 1080
QY 1081 FYLAAMIRILIEILITSGRTVRYNSGLPDBEAMLEAKINOVPAATVVSADWTEFDTAH 1140
DB 1081 FYLAAMIRILIEILITSGRTVRYNSGLPDBEAMLEAKINOVPAATVVSADWTEFDTAH 1140
QY 1141 NNTSELPAALIERIGTPAAAVNLPERRCGRTTLAKLGSVENUGLLDSGAAMPCCNT 1200
DB 1141 NNTSELPAALIERIGTPAAAVNLPERRCGRTTLAKLGSVENUGLLDSGAAMPCCNT 1200
QY 1201 IFSAVMTLTFGVGFAPFKGDSLLCGSHYLRFPASRLHMERKTKHLKVEYOKIYVY 1260
DB 1201 IFSAVMTLTFGVGFAPFKGDSLLCGSHYLRFPASRLHMERKTKHLKVEYOKIYVY 1260
QY 1261 IGLVSAQVVLDPVRSALKIFGRCTSELLYSKYVEAVRDTTKGMSDARYSLLCHMSA 1320
DB 1261 IGLVSAQVVLDPVRSALKIFGRCTSELLYSKYVEAVRDTTKGMSDARYSLLCHMSA 1320
QY 1321 CYNNYAPSAAYITIDAVVRPGRGPRPQOLVVRHQAAPAYSTYPAVPAASCLDHYF 1380
DB 1321 CYNNYAPSAAYITIDAVVRPGRGPRPQOLVVRHQAAPAYSTYPAVPAASCLDHYF 1380
QY 1381 EPRQAAAPAGFVATCAKETPSSSLTAKAGVATTSHTVATGTAPPSPPMDAPANGSEIL 1440
DB 1381 EPRQAAAPAGFVATCAKETPSSSLTAKAGVATTSHTVATGTAPPSPPMDAPANGSEIL 1440
QY 1441 TPETPSTSSPSSSSSDSSTSCGRSLSGDTRATTEDINSRKPQDRQSRSSBCLDRSG 1500
DB 1441 TPETPSTSSPSSSSSDSSTSCGRSLSGDTRATTEDINSRKPQDRQSRSSBCLDRSG 1500
QY 1501 ERTGSSLTAPTPASPSFSPSEBAALATGPTVAAAASBPATBSCANDVQAARTTPPAPPL 1560
DB 1501 ERTGSSLTAPTPASPSFSPSEBAALATGPTVAAAASBPATBSCANDVQAARTTPPAPPL 1560
QY 1561 GSQSAVASKPYRPTTARKKEVTPLAHMKGVYGRPREYBSPETAAVYQALISGRYPQK 1620
DB 1561 GSQSAVASKPYRPTTARKKEVTPLAHMKGVYGRPREYBSPETAAVYQALISGRYPQK 1620
QY 1621 TULSSDASKGYRTKCSQSTSFPAASADYQARDQTVAVCAAEMASCIHEPLASSA 1680
DB 1621 TULSSDASKGYRTKCSQSTSFPAASADYQARDQTVAVCAAEMASCIHEPLASSA 1680
QY 1681 ASADIKRIRSTSDSVDPYKISKA 1704
DB 1681 ASADIKRIRSTSDSVDPYKISKA 1704

RESULT 2
US-09-194-613-5
Sequence 5, Application US/09194613
Patent No. 6251654
GENERAL INFORMATION:
APPLICANT: GORDON, Karl H.
APPLICANT: HANZLIK, Terry N.
TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDermott, Will & Emery
STREET: 600 13th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3096
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194.613
FILING DATE: 30-NOV-1998
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: Bucca Ph.D., Daniel
 REGISTRATION NUMBER: 42,368
 REFERENCE/DOCKET NUMBER: 50179-061
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-756-8000
 TELEFAX: 202-756-8087
 INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1233 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-194-613-5

Query Match 16.4%; Score 1444.5; DB 3; Length 1233;
 Best Local Similarity 31.0%; Pred. No. 6,8e-119;
 Matches 425; Conservative 195; Mismatches 525; Indels 227; Gaps 44;

QY 5 ATDVARYAADAAYANVLOQRAVKLDPAAPLKALETHRLYYPRKGGTLPTQPHIL 64
 DB 4 ASKQLRLDAQERAKA-----AFQIDFIASVETLEDAQEKYEGMMFRSGTKLPSTRIKL 57
 QY 65 AGHQAEEVTLNHPARGSTVLEIGPSLSALKLHGAPNAPVADYHCGCTKYGTSDGSRH- 123
 DB 58 AIDLRVAEKDLRHHVKNVPTVLEIGPSVESV--RIAVQTDKERVHCGCTSDARDNLHFN 115
 QY 124 ---ITLESRSVATGRPEFRADSLANGIASRTFCVDGVGSCAFKSRVGIANHSLY--- 177
 DB 116 KIGVEAHYDRKIG-----PDALLAAGIPTDTFCVDGFSNCGYQPLAIGHSLVDPG 168
 QY 178 -DVTLEELANAFENGILMTVRAFMAMEPELLYMDNVNAELGTRFHVIEEPMAYKCAFO 236
 DB 169 ESNSIMDVAKGMLHGHVYVAMWHLVLELLTLDADNIEGYSI----- 213
 QY 237 GGLRLHFPFLDFINESQERIERLARGSYSRAVIFSGDDMDGAYLDFHTWLAYLL 296
 DB 214 -----RPEETGALPCTKRRKALFSGYNDGSAYVVDHAWMACMLK 253
 QY 297 VRNVPFPGSLNIEVORRHGSSIEILRTAPRPGDRLMAVPRISQGLCHIPNI--FYVA 354
 DB 254 HRGVDPFYGSILIDIOGRFGMHKLIKITHGSSGSLTTFPLSKGLIWPNI VKIMYP 313
 QY 355 DASGTEHTILTSQHKVNMMLNFMQTPREK--ELVDMTVLMSFAPARLRAIVASEVTES 412
 DB 314 KAK-HEBEYIVTDKKYEGCVVYGVTRVQSSGKITLAIYQVIRTLRIILNGTVHEK 372
 QY 413 SMNISPADLVRTVSVLYVLIHIERRAAVAKTKADVPGETSWESLKVYLSGCCGLRN 472
 DB 373 TWITAEQDIERLAVSIIFRRKVERAVSEKALMRQKK-----CKSAE 414
 QY 473 LKGTDVVFTKRVV---DKYR-----VHSLGDIICDV-RLSP--EQ 506
 DB 415 KQALLPVMKRIAMFQDKQIDBEVVRKYLECLKAQPMIHAKVYVNCETKRINPVAE 474
 QY 507 VGPLSKVRPARVFNDEEELVLEACQYNERPVSPVEEPQGFADIMHATPAASLPE 566
 DB 475 VG-----PKNHLATATGLRELQRE-----IPSNAPEDPGAKAMHSHADLDI 517
 QY 567 YRATLQAGLNTD-----VKQKITLENALKTIIDGLTSPVNGLEM--YEGPFGSGKT 616
 DB 518 Y----AEGRLDSAKERAAGKQSLATILQAFQVLGKTGCEGCNNIIEIYWTGPGSGKS 573
 QY 617 GTLLALEAAGKALYVAPTRLEBAMDRRIKPPASATQHVALLARATABGAPATV 676
 DB 574 RAKKPRADLQGGTLVYCAPTRTLRDALDESIVHRSRCTYHNAHVAKEG-NRPDVI 632
 QY 677 VIDECFNPVLYVAIVHALSPSSRIVLGVVHIGITIDPGSTANMPLVRDVVKQCRRT 736
 DB 633 VIDEAETTPACYGVQTMHASPSRSRIIVCLGDPHQIGYIDFSRDKDLKPFSSIIAECRTR 692
 QY 737 FNGTKRPADVVAATFFQSLYPGCTTSGCVASISHA---PDVRNQAQTLCTQDEKS 793

DB 693 FNTTYRQPDVNLPIPKLYPDAISPSKQLSIRVLTARASVTRTHAQLT--TQDPK 750
 QY 794 RHQAGAMTVHBAQGTTFASVILHNGSTAEOQLAEKSHLVIGITRTHNLVTRDPTGD 853
 DB 751 H--SEBPVTAHBPQAR-TDVIYHAGVTLPERALLKEKVRHINVALTHTNALYLRD---- 803
 QY 854 IERQLNHSAAEYFTDIPAPLBITVK--PSEB-----VQRNEMATIPPSAIPHG 903
 DB 804 -----BSEKGEVPSLMTTPSSWYTRCTVDQMPDPVAVAREN--GSGPDSHHIG 855
 QY 904 AIHLRNKFGDQDCCVALAKTGYEVFGRAKINVELAPDAPPKPRH----- 952
 DB 856 AITLIGE-LGKLD-----TGQVYF-----ESEAVPTHRVVLIDGNLDSG 896
 QY 953 -----AFQ-BGVQWVKTNAHNGALQTLISRYTRASDLPHEAKEDVYKMLNSIDRH 1006
 DB 897 PDREYPMQFNLIGTKYTNIKDQQLHLTVGHYARKINS-----SSRETPSLTRESQA 952
 QY 1007 WDMVTYEDADRAV---FETQLKFTQRCGTVEEDLLEBDDPYIND---IDPLMTQOKV-S 1059
 DB 953 QEWIPRHABPEQVDCSFATPCRCRGRGHGVDDI---EDFWSNEGQRISTHLGQOQKMD 1009
 QY 1060 PKPINTGKVGQIAHASKSLNPLVLAAMIRILBEILRTGSRTRYSNGLPDEEAMLEAK 1119
 DB 1010 PTKL---KLGQISAEKCNIALSAMVRLIQDMGSEKFT-PANGQSRDMSITIEAR 1065
 QY 1120 INQVPATFVSADWTEPDTHANNTSBELPALLERIGTPAAAVNLFREKCGKTLRAKGL 1179
 DB 1066 L-QEKAREFSIDIKEDDTVHNVWSILVFSMRCDR-GCPHLLIEYFEKRSKSLTSR-I 1122
 QY 1180 GSYEVGDLDSGAAMPCCRTIFSAVMNLTLPFGYVPAAKGSDSLSCGHYLRPDSRL 1239
 DB 1123 GSYDVSPMDSGAVMTIARTLTLPASGLMLALFVGVPPIAKGDDVFLAANN-LYLDARL 1181
 QY 1240 HMGERYTKLKYEVQKIYPYIGLVSABQVVDPVYSALKIFRCRYTSEL 1291
 DB 1182 RMGSYLAANNLKIETRAVSPFIFYSQAAVTADVRLAIRTYGRSITKAVMI 1233

RESULT 3

US-09-462-606-2
 Sequence 2, Application US/09462606
 Patent No. 6432408

GENERAL INFORMATION:
 APPLICANT: MENG, XIANG-JIN
 APPLICANT: Emerson, Suzanne U.
 APPLICANT: Purcell, Robert H.
 TITLE OF INVENTION: A SWINE HEPATITIS B VIRUS AND USES THEREOF
 FILE REFERENCE: 20264267US1
 CURRENT APPLICATION NUMBER: US/09/462,606
 PRIOR FILING DATE: 2000-06-12
 PRIOR APPLICATION NUMBER: US 60/053069
 PRIOR FILING DATE: 1997-07-18
 PRIOR APPLICATION NUMBER: PCT/US98/14665
 NUMBER OF SEQ ID NOS: 65
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2
 LENGTH: 1708
 TYPE: PRT
 ORGANISM: Hepatitis B virus
 US-09-462-606-2

Query Match 5.8%; Score 510.5; DB 4; Length 1708;
 Best Local Similarity 21.3%; Pred. No. 2.2e-35;
 Matches 389; Conservative 183; Mismatches 558; Indels 699; Gaps 81;

QY 13 AADVAYANVLOQRAVKLDPAAPLKALET--LRLYYP--LPRKGGTLPTQPHILAGHQ 68
 DB 22 AANSALANAVVR-----PFLSRVQTEILINMQPRLVFRPEVL--WNHPI-----Q 68
 QY 69 RVAEVTLNHPARGRS-TVLEIGPSLSALKLHGAPNAPVADYHCG-TRKYGTSDGSRHITA 126

Db 69 RAHNELEBOYCRARAGCCELEVG--AHPRE-INDNENV---LHRCFLRPVGRDVORWYSA 121
 Qy 127 LE-----SRGVAAGREPEFKADASLLANGIASRTPFCUDVGSCAFKSVAGLANSLVDV 179
 Db 122 PTRGPANCRSALRGLRP-----VDRTYCFDGFSCRCAFAAETGVALYSLHDL 169
 Qy 180 TLEELANFENHGLHVAFMHMPBELLYMNVNVAELGYRPHVIEEPMAVDCAFQGGD 239
 Db 170 MPADVAEMARHGMTRLYAHLHLEBVLPRCTYHTTSLTH-----DGD 215
 Qy 240 LRLHPELDFINESQERRIERLAAGSYRRRAVIFSGDDWDGDAYLHDFHTWLYL--- 295
 Db 216 -----RAVVEGDTAG--YXHDVSLRAMIRTK 244
 Qy 236 LYRNPTRPFGSLHIEVORRHSSIELRTRAP-----PGDKMLAVPTSGCLRI 347
 Db 245 IVDGHP-----LVIERVNAIGCHFVLTLTAAPBPSMPYVYPRSTEVYVSIIPGGS 298
 Qy 348 PNIIFYADASGTEHHTILTSOKVMMLNFMQTRPEKELVDMTVLMSFARARLRAIVNAS 407
 Db 299 PSLF--PSACSTKSTFHAVPVHIMDRMLFGATLDDQAF-C-SRLMTYLRGISYKVTGA 355
 Qy 408 EYTESMNI SPADLVRTVSLVYLHIIBERRAAVAV----- 443
 Db 356 LVANCGMNASBDALTAVALTAVALTICHOYLRTOAISKGMRLVEYHAKFITRILYSLF 415
 Qy 444 -KTAODV-----FGETSFWES-----LKHVUGS--CCGLR 471
 Db 416 EKSGRDYIPGRQLOFYAOCRRWLSAGFHLDPVLVDESVPORCRFTFLKVKVGCPCFMR 475
 Qy 472 -----NLKGTDVVTKVVNDKVRV- 490
 Db 476 MLGECTCFLEABGLVGDYGHNDNAYBSSEVDPAEPALDVSCTYAAHGRQLEALYRL 535
 Qy 491 ----- 490
 Db 536 NVPHDIAARSLTATVELTASPDRLCRTLGNKTRTTVDGAMLEANGPEQYVLSFD 595
 Qy 491 -----HSL-----GDI--ICD----- 499
 Db 596 ASROGMAOSHSLTYELTPAGLOVRISNGLDCTATPPGGA PSAAPGVAFCALYRY 655
 Qy 500 -----VRLSPEOV-----GF----- 509
 Db 656 NRFTORHSLTGSLMHPBGLGIPPPSPGHIWEPANPCGEGTLTYRTWSTSGSSDPS 715
 Qy 510 -----LPSRVPARVPHRELEVLREAGCYNERPVSTPVEPOG----- 551
 Db 716 PPEAAAPVLAAPGLPHTRPV-----SDIWLPPPSKESQVDAASVPAPREPAGLPS 769
 Qy 552 -----PDADL-W----- 557
 Db 770 IVLTLRPLRPVKRPPTRPRSRRLLYTPDGAKYASLSLSPSCNMVLANSPGHRPG 829
 Qy 558 -----HATAASLDEYRA-----TLQAG- 574
 Db 830 GGLCHAFORPREAAYPTFEFIRREGLAATYTLTPRIHVA--PDYRVEONPKRELAAY 886
 Qy 575 -----LNTDVKQLKITL----- 586
 Db 887 RETCSRGTAAVPLGSGIYQVPSLSFDMERNHRPGDELVLTPBAAMFEANKPAQPA 946
 Qy 587 -----ENAKLT-----ID-----GLTSPVRGLEMYE--GPPOSGTGTLLAAL 623
 Db 947 LTTEDTARTANLALEIDPATDVGRACACTISP-GIVHYOFTGTVGSPSGKRSI----- 1000
 Qy 624 BAAGKALYVAPTRELEAMDRIRKIPPSATONVA--LAILRATAGAPATVVIDEC 681
 Db 1001 -QGGVDVVVVTRELRNSWRR--GPAFTPHTRATYTKR-----VVIDEA 1046
 Qy 682 FMFPLVVAIVHALSPSSRIYVGVHVGIGIDFOGTSAMBLVDVVKOCRRRTFNOTK 741
 Db 1047 PSLP-PHLLH-MQRASSVHLGDPOQPAIDFHAQL-VPAIRP--ELAPTSMWHVTH 1101

Qy 742 RCPADVATTFPQSLYPCCTTSGCVASISHVAPDYRNSOACTLCFTQBEKSRHAGBGM 801
 Db 1102 RCPADV--CELIRGVYPIQTTSRVLRSLFWNER---AIOQKLVFTAAKAA--PGAI 1153
 Qy 802 TVHEAGRTFASVILHNVGASTABOGLABKS--HLLVGITRTHNLTYRDPFGDIEROLN 859
 Db 1154 TVHEAGATFETITII---ATRDARGLQSSRAHIVALTHTTEKCVILLDARG-LIREVG 1209
 Qy 860 HS-----AKAEVFTDIPAPLEITTVKSEBEVORN-EVMATIPQSATPHGAIHILR 909
 Db 1210 ISDVIVNNFFLAGSGVGHRRP-----SVIRGNPDQMLGTLOAFPPSCQI--SAVHQLA 1261
 Qy 910 KNFGDPOCCGVAALAKTGVYVPGCAKINVELABEADAPKRRARQ-BGVQVQKTMASN 968
 Db 1262 BELGRP-APVAALVPCPELBQGLLYNPOELTVSDV---LVBELTDIYHCRMAASQ 1316
 Qy 969 KHALQTLISRTKRSADLPHE--AKEDVKMNLSDRHMDPWTVEDARAVFETOLKF 1027
 Db 1317 KRAVSTLVGRYGRRT---KLYEAAHSDVR--BSLAFTPIPGVQATTGELYELVZAM 1370
 Qy 1028 TORGGTEDLLEPD--DPYIRIDIFLMKTOOK-VSPKPIITGVQOGIAHASKSLNFYLA 1084
 Db 1371 VEKQDGSVAVLELDLNDVSRITFPQKCNKFTGTETIANGKVGQGISAMSKTCALFG 1430
 Qy 1085 AMIRILE-EILRTSGRTYRISNGLPDEBEAMLEKINQVPHATVSDMTPEFTAHNT 1143
 Db 1431 PMPRAIERELIALLPNNIFYG---DAYBSVFAAASGASCMVFENDPSEFSDTONNF 1486
 Qy 1144 SELFALLERIGTPAALVNLFRERCGKRTLA-----KGLGSVEVDGLDGAAMTPCR 1198
 Db 1487 SLGLECVMBEGMOWMILRLYLVRSAWILQAPRESLKGPWKH-----SGEGTLLM 1540
 Qy 1199 NTIFSAAVMLTL--FRGVKFAAFKGDG--LTCGSHYLPDASRLMGRYKTKLKVYEQ 1255
 Db 1541 NTVMMAIILHCEYFRDPRVAAPFKGDDSVILCSDYRQSRNNAALJAGGLK--LKVDYR 1597
 Qy 1256 KIVPIYIGLLVSAEOVVL-DPVSALAKIFGRCTSSLLSKYVAVRDTTKGSDARHSL 1314
 Db 1598 PGLAGVAVAPGLCTLPDVVFA---GR--LSB-----KMWGP----- 1631
 Qy 1315 LCHMSACYNNAPEESAAYIIDAIVRFRG 1343
 Db 1632 -----GPERABQLRAVCDPLRG 1649

RESULT 4
 US-08-840-316-1
 ; Sequence 1, Application US/08840316
 ; Patent No. 6054567
 ; GENERAL INFORMATION:
 ; APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
 ; APPLICANT: Tearev, Sergei, A., and Robinson, Robin A.
 ; TITLE OF INVENTION: Recombinant Proteins Of
 ; TITLE OF INVENTION: A Pakistani Strain Of Hepatitis B And Their
 ; NUMBER OF INVENTION: Use In Diagnostic Methods And Vaccines
 ; NUMBER OF SEQUENCES: 111
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: MORGAN & FINNEGAN
 ; STREET: 345 PARK AVENUE
 ; CITY: NEW YORK
 ; STATE: NEW YORK
 ; COUNTRY: USA
 ; ZIP: 10154
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: FLOPPY DISK
 ; COMPUTER: IBM PC COMPATIBLE
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: WORDPERFECT 5.1
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/840,316
 ; FILING DATE: 11-APR-1997
 ; CLASSIFICATION: 424

1 PRIOR APPLICATION DATA:
 2 APPLICATION NUMBER:
 3 FILING DATE:
 4 CLASSIFICATION: 424
 5 ATTORNEY/AGENT INFORMATION:
 6 NAME: Richard W. Bork
 7 REGISTRATION NUMBER: 36,459
 8 REFERENCE/DOCKET NUMBER: 2026-4255
 9 TELECOMMUNICATION INFORMATION:
 10 TELEPHONE: (212) 758-4800
 11 TELEFAX: (212) 751-6849
 12 INFORMATION FOR SEQ ID NO: 1:
 13 SEQUENCE CHARACTERISTICS:
 14 LENGTH: 1693 AMINO ACID RESIDUES
 15 TYPE: AMINO ACID
 16 STRANDEDNESS: UNKNOWN
 17 TOPOLOGY: UNKNOWN
 18
 19 IS-08-840-316-1

Query Match	5.7%;	Score 500;	DB 3;	Length 1693;
Best Local Similarity	21.2%;	Pred. No. 1.9e-34;		
Matches 371;	Conservative 166;	Mismatches 575;	Indels 614;	Gaps 72

QY	2	YAKATDVAVVAAADVAANVYVLOQRAVKUDFAPRLKALETHRLYYP--LRFEGGTLPP	55
Db	6	FIKAGITTALEQAAALAAANSALANAVVRPFUSHQOIELLNLWMPROUVRPEVF--W	63
QY	60	QHPILAGHORVAEEVLHNEFARGRS--TVLEIGPSLHALKLHGAPNAVVDYHGC--TKYGT	117
Db	64	NHPI-----QRVHNELELYCRARSGRCLTGAHPRS---INDNPV---YHRCPLRPAG	112
QY	118	RDGSRHITALE-----SRSAVATGRPEFKADASLLANGIASRTFCVDGVGSCAFKSRVG	170
Db	113	RDVGWYTAPTFRGPANCRRSALRGLP-----AADRTYCFDGGSGCNFPAETG	160
QY	171	IANSLSYDVTLEBELNAPFENHGLHVRAPFMHPPEELLVMDNVYNAELGIRFHVITEPMAY	230
Db	161	IALSYLHOMSPSDVAEMFRHGMTRLYAALHLPPEVTL-----PGTY	203
QY	231	KDCAFOGDDLRAHPPELDFINESQERIERLARGSYSRRAVYFSGDDWDGAYLHD--	287
Db	204	RTAST---LLIH-----DGRV-----VTYEGDTSAG--YHGDVSN	235
QY	288	FHTWLAVLLVRNYPPTPFGFSLHIEVORRHGSSIELRITRAP-----PGDRMLAVPR	339
Db	236	LRSW-----IRTKVTGDRPLVIERBRAIGCHFVLLTLTAAPESPMYPYPRSTEVYR	290
QY	340	TSQGLCRIPNIFYVADASTEHKTLTISOHKVNLMLNFMOTRPEKELVDMTVMSPARAR	399
Db	291	SIFGGGSPSLF--PTSCSTKSTFHAVPAHIMDLMLFGATLDDQAF-C--SRMLTYLRGI	347
QY	400	LRAIVVASEVTESSNNISPADLVRTVLSLYV-----LHIEERRRAAVAV	443
Db	348	SYKVTVGLVANEGMNASDALTAVTAAVLYTICHORYLRTQAIISKGRMLREHEHOKFI	407
QY	444	-----KTAKD-----DVFGETSEWESL-----KHV	463
Db	408	TRLYSMLPEKSGRDYIPGRQLEFYAQCRMLLSAGFLHDLRVLVFDSEAPCHCCTAIRKAV	467
QY	464	LGSCC-----GLRLKGTDVVFTKR	483
Db	468	SKFCFEMKMLGOECTCFLQPAEGVVDQGHNDNEAVGSDVDPAESAISDISGYYVPGTA	527
QY	484	VVDKRYVSL-----GDIIIDVYL-----	502
Db	528	LQPLQALDLPRAEIVARAGRLTATYKVSQVDGRIDETLLGNKTFRTSPVDGAVLETTNGP	587
QY	503	-----	502
Db	588	ERHNLSFDASOSTMAAGPESLTYAASAAGLEVRYYAAGLDHRAVFAFGVSPRAAPGEVTA	647
QY	503	-----	502

D	648	FC\$ALYFENBAQSL\$TGNFWHREBOLBPAPF\$SGHWT\$E\$ANPFC\$G\$STLYTTRT\$E	707
Q	503	-----SPBQ---VGF-----LPSRV-----PPA-----	517
D	708	VDAP\$P\$AQ\$PDLG\$T\$P\$S\$P\$P\$R\$A\$T\$P\$A\$P\$L\$P\$P\$A\$D\$P\$S\$L\$A\$R\$B\$A\$P\$A\$G\$A\$T\$A\$R\$A	767
Q	518	-----	517
D	768	PAITHQ\$T\$AR\$H\$R\$E\$L\$F\$T\$Y\$P\$D\$S\$K\$V\$F\$A\$G\$L\$F\$E\$S\$T\$C\$T\$W\$L\$V\$N\$A\$S\$V\$D\$H\$R\$P\$G\$G\$L\$C\$H\$A\$F\$O\$R\$Y\$P\$A	827
Q	518	-----RV\$H\$D\$R\$E\$B\$L\$E\$V\$L\$R\$E\$A\$C\$Y\$N\$-----E\$R\$E\$V\$P\$S	542
D	828	SFDA\$A\$F\$V\$M\$D\$G\$A\$A\$Y\$T\$L\$T\$P\$R\$P\$I\$H\$A\$P\$D\$Y\$R\$E\$H\$N\$K\$R\$E\$A\$V\$R\$E\$C\$S\$L\$G\$T\$A\$Y\$P\$D\$LL\$G	887
Q	543	TP\$V\$E\$E\$P\$O\$G\$F\$D\$A\$D\$L\$M\$H\$-----A\$T\$A\$S\$L\$P\$E\$Y\$R\$A\$-----T\$L\$O\$G\$N\$T\$D\$Y\$K\$--L\$K	583
D	888	T\$G\$I\$O\$V\$E\$I\$G\$S\$F\$D\$M\$E\$N\$H\$R\$O\$D\$E\$L\$Y\$P\$E\$L\$A\$R\$F\$E\$A\$N\$R\$E\$P\$T\$C\$P\$L\$T\$--I\$T\$E\$D\$V\$A\$R\$T\$A\$N\$L\$A	945
Q	584	I\$T\$E\$N\$A\$L\$---K\$T\$I\$D\$G\$L\$T\$L\$S\$P\$G\$L\$E\$W\$E\$---G\$P\$G\$S\$G\$T\$O\$T\$L\$A\$A\$E\$A\$G\$K\$L\$Y\$V\$A\$P\$T\$R	637
D	946	I\$E\$L\$S\$A\$D\$V\$G\$A\$C\$G\$C\$V\$T\$P\$--G\$V\$O\$Y\$O\$F\$T\$A\$G\$V\$P\$G\$S\$K\$S\$T\$O\$A\$-----D\$V\$D\$V\$V\$P\$T\$R	998
Q	638	E\$L\$R\$E\$A\$M\$D\$R\$I\$K\$P\$S\$A\$T\$O\$H\$A\$L\$A\$I\$L\$T\$--R\$E\$A\$T\$E\$G\$A\$P\$A\$T\$V\$I\$D\$E\$C\$H\$F\$P\$L\$Y\$V\$A\$I\$V\$H\$A\$L	695
D	999	E\$L\$R\$A\$M\$R\$R\$---G\$P\$A\$F\$T\$P\$T\$H\$A\$V\$T\$G\$R\$-----V\$I\$D\$E\$A\$P\$S\$L\$P\$--P\$H\$I\$L\$L\$H\$--M	1043
Q	696	S\$P\$S\$R\$I\$V\$L\$V\$D\$V\$H\$O\$I\$G\$I\$F\$D\$O\$G\$S\$A\$N\$P\$L\$Y\$R\$D\$V\$Y\$K\$O\$C\$R\$E\$T\$F\$N\$O\$T\$K\$C\$P\$A\$D\$Y\$A\$T\$T\$P\$O\$S	755
D	1044	Q\$R\$A\$T\$V\$H\$L\$G\$P\$N\$O\$I\$P\$A\$I\$D\$E\$H\$A\$G\$L\$--V\$P\$A\$I\$R\$P\$D\$L\$--A\$P\$T\$S\$M\$H\$Y\$T\$H\$R\$C\$P\$A\$D\$V\$--C\$E\$L\$I\$R\$G	1098
Q	756	L\$Y\$P\$C\$T\$T\$G\$C\$V\$A\$S\$I\$S\$H\$A\$P\$D\$Y\$R\$S\$A\$O\$T\$C\$F\$T\$O\$E\$E\$K\$S\$R\$G\$A\$E\$G\$A\$M\$V\$H\$E\$A\$O\$G\$T\$P\$A\$S\$V\$I	815
D	1099	A\$Y\$P\$H\$I\$O\$T\$T\$S\$R\$L\$R\$E\$L\$F\$W\$G\$E\$R\$-----A\$V\$O\$K\$L\$V\$T\$O\$A\$A\$A\$N\$--P\$C\$S\$V\$Y\$H\$E\$O\$G\$T\$Y\$E\$T\$T	1152
Q	816	L\$H\$N\$G\$S\$T\$A\$E\$O\$K\$L\$A\$E\$K\$S\$--H\$L\$V\$G\$I\$T\$H\$Y\$T\$H\$N\$L\$Y\$I\$R\$D\$T\$G\$D\$I\$E\$R\$O\$L\$N\$S\$A\$A\$E\$Y\$T\$D\$I\$P\$A\$P	873
D	1153	I\$I\$---A\$T\$A\$D\$A\$R\$G\$L\$I\$O\$S\$R\$A\$H\$A\$I\$V\$A\$L\$T\$H\$T\$E\$K\$C\$V\$I\$D\$A\$P\$G\$--L\$R\$E\$V\$G\$I\$S\$--D\$A\$I\$V\$N\$P\$F\$I\$A\$G	1207
Q	874	L\$E\$I\$T\$T\$V\$P\$S\$E\$V\$O\$R\$N\$-----E\$W\$A\$T\$P\$P\$O\$A\$T\$P\$H\$G\$A\$I\$H\$L\$R\$K\$N\$G\$D\$P\$D\$C\$G\$C\$V\$A\$L\$A\$T\$G\$Y	928
D	1208	G\$E\$I\$G\$H\$Q\$R\$P\$S\$V\$P\$R\$A\$N\$P\$A\$N\$V\$D\$T\$L\$A\$P\$P\$S\$C\$E\$I\$--S\$A\$F\$H\$E\$L\$A\$B\$E\$L\$G\$H\$R\$P\$--A\$P\$V\$A\$V\$L\$P\$C\$P	1264
Q	929	E\$V\$F\$G\$R\$A\$K\$I\$N\$V\$E\$L\$A\$P\$D\$P\$P\$K\$P\$H\$A\$P\$O\$--E\$G\$O\$V\$W\$K\$V\$T\$A\$N\$K\$H\$Q\$A\$L\$O\$T\$L\$S\$R\$Y\$K\$S\$A\$D\$L	987
D	1265	E\$L\$B\$G\$L\$Y\$L\$P\$E\$L\$U\$T\$C\$D\$S\$V\$---T\$E\$F\$L\$D\$I\$O\$Y\$H\$O\$R\$M\$A\$P\$B\$O\$R\$K\$A\$V\$S\$T\$I\$V\$G\$R\$Y\$O\$R\$T\$---	1317
Q	988	P\$L\$E\$A\$K\$--E\$D\$Y\$K\$M\$L\$N\$S\$I\$D\$R\$H\$M\$D\$T\$V\$T\$E\$D\$A\$D\$R\$A\$V\$E\$T\$O\$L\$K\$F\$O\$R\$G\$T\$V\$E\$D\$L\$E\$P\$D\$--D\$P\$Y	1044
D	1318	K\$L\$V\$N\$A\$S\$H\$S\$D\$V\$R\$---D\$S\$L\$A\$R\$P\$I\$A\$G\$P\$O\$V\$Y\$T\$C\$E\$L\$Y\$E\$L\$E\$M\$A\$V\$E\$K\$G\$O\$G\$S\$A\$V\$L\$E\$L\$D\$L\$C\$R\$D	1374
Q	1045	I\$R\$D\$I\$D\$F\$L\$M\$K\$T\$O\$Q\$K\$--V\$S\$P\$K\$I\$N\$T\$G\$V\$G\$G\$I\$A\$H\$S\$K\$L\$M\$F\$V\$A\$A\$V\$R\$I\$E\$B\$--I\$T\$R\$G\$S\$T\$V\$R	1102
D	1375	V\$S\$R\$I\$T\$F\$O\$K\$O\$N\$K\$T\$T\$E\$T\$I\$A\$H\$K\$V\$G\$G\$I\$S\$M\$S\$T\$P\$C\$A\$L\$G\$P\$M\$P\$R\$A\$I\$E\$K\$A\$L\$I\$A\$L\$L\$P\$O\$G\$V\$F	1434
Q	1103	Y\$S\$N\$G\$L\$P\$D\$E\$B\$--E\$A\$M\$L\$E\$A\$K\$I\$N\$O\$V\$P\$A\$T\$E\$V\$S\$A\$D\$W\$T\$E\$F\$D\$T\$A\$H\$N\$T\$S\$E\$L\$E\$P\$A\$L\$L\$E\$R\$I\$G\$T\$P\$A\$A	1161
D	1435	Y\$G\$D\$A\$F\$D\$T\$V\$F\$A\$A\$V\$A\$A\$K\$A\$S\$V\$---F\$E\$N\$D\$F\$S\$F\$D\$T\$O\$N\$N\$F\$S\$L\$G\$A\$L\$C\$A\$L\$E\$B\$G\$M\$P\$O\$M\$L	1489
Q	1162	V\$N\$L\$R\$E\$R\$O\$C\$K\$R\$I\$L\$A\$K\$G\$L\$S\$Y\$V\$D\$G\$L\$-----L\$D\$S\$G\$A\$M\$T\$P\$C\$R\$N\$T\$I\$F\$S\$A\$A\$V\$M\$L\$T\$--F\$R\$V	1214
D	1490	I\$R\$L\$V\$-----H\$L\$I\$S\$A\$M\$I\$T\$O\$A\$P\$K\$E\$S\$L\$G\$F\$M\$K\$K\$S\$G\$B\$G\$T\$L\$M\$N\$T\$V\$M\$M\$A\$Y\$T\$H\$C\$Y\$D\$R\$D\$L	1543
Q	1215	K\$E\$A\$F\$K\$G\$D\$S\$L\$--L\$G\$S\$H\$Y\$L\$R\$F\$D\$A\$S\$R\$L\$M\$G\$E\$R\$Y\$T\$K\$H\$L\$K\$A\$V\$E\$O\$K\$I\$V\$P\$Y\$I\$G\$L\$V\$S\$A\$E\$Q\$V\$L\$--	1272
D	1544	Q\$V\$A\$F\$K\$D\$D\$S\$T\$V\$L\$C\$E\$S\$E\$V\$R\$O\$B\$G\$A\$V\$L\$A\$G\$G\$K\$---L\$K\$V\$D\$P\$R\$I\$G\$L\$A\$G\$V\$V\$A\$G\$D\$A\$L\$P	1600
Q	1273	D\$P\$V\$S\$A	1278
D	1601	D\$V\$A\$P\$A	1606

RESULT 5
US-08-809-523-1

```

: Sequence 1, Application US/08809523
: Patent No. 6207416
: GENERAL INFORMATION:
: APPLICANT: Tsarev, Sergei. A., Emerson,
: APPLICANT: Suzanne U., Purcell, Robert H.
: TITLE OF INVENTION: Recombinant Proteins Of
: TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
: NUMBER OF SEQUENCES: 107
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/809,523
: FILING DATE: 28-MAY-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/US95/13102
: FILING DATE: 03-OCT-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US08/316,765
: FILING DATE: 03-OCT-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/947,263
: FILING DATE: 18-SEP-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Richard W. Bork
: REGISTRATION NUMBER: 36,459
: REFERENCE/DOCKET NUMBER: 2026-4032US4
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 758-4800
: TELEFAX: (212) 751-6849
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1693 AMINO ACID RESIDUES
: TYPE: AMINO ACID
: STRANDEDNESS: UNKNOWN
: TOPOLOGY: UNKNOWN
: US-08-809-523-1

```

```

: Query Match 5.7%; Score 500; DB 3; Length 1693;
: Best Local Similarity 21.2%; Pred. No. 1.9e-34;
: Matches 371; Conservative 166; Mismatches 575; Indels 614; Gaps 72;

```

Qy	2	YAKATDVAVVAADVAVVAVLQQRVAKLDFAPPLKALETLRLYYV--LRFKGGTLPT 59
Db	6	FIKARGITTAIEQALAAANSALANAVVVRPFLSHQIILINLMQPRQLVRPVEV--W 63
Qy	60	QHPILAGHVRVAEVLVHNAFGRS-TVLEIGPSLHSAALKHAGAPNAVVDHYGC-TKYGT 117
Db	64	NHPI---QRVIHNELELYCRARSGRCLTIGAPRS---INDNPV---VHRCFLPAG 112
Qy	118	RDGSRHITALE-----SRSVATGRPEFKADASLLANGIARSTPCVDVGSCAFKSRNG 170
Db	113	RIVQKWTYAPTRGPANCRSALRGLP-----ADDTTCFDGFGGCFPAETG 160
Qy	171	IANHSLVDTLELNAFENHGLHNVRAFMEHPEBELLYNDVNVNAELGYRFVIBEPNAV 230
Db	161	ILVSLHDMSPSDVAEAMFRHGWTRLYALHLPEVLL-----PGTY 203
Qy	231	KCAACQGGDLRLHFPBLDINSQERIRIARLARGSYRAVIFSGDDWDGAYLHD--- 287
Db	204	KFASY---LTH-----DGRV-----VYTEGDSAG--YNHDSN 235
Qy	288	FTTWLAVLVVRYPTFPFGSLHIEVQRHGGSSIELRTRAP-----PGDRMLAVVR 339

```

: Db 236 LRSW-----IRTKVTDGHPVIERVVALGCHPULLTAABEPSPMPVVPFRSTREYVR 290
: Qy 340 TSQGLCRIPNIPTVADASGTEHKTILTSQHKVMNLNFMQTRPEKELVDMTVLSFAPAR 399
: Db 291 SIFPGGCTPSLF--PTSCSTSTHVAVAHIMDBLMFGATLDDQAF-CRLMTYLRGI 347
: Qy 400 LRAIVVSEVYBESSMNISPADVTVVSLV-----LHIERRAVAV 443
: Db 348 SYKTVGLVANEGNNSBDALTAVITTAAYLTICHQRYLTQALSKGRRLREHAKPI 407
: Qy 444 -----KTAKD-----DVEGETSFMSL-----KHV 463
: Db 408 TRLSWLFKSGRDIYIPROGLEFYAQCGRWLSAGHLDPRVLVVDDESAPCHRTAIRAV 467
: Qy 464 LGSC-----GLNMLKGTDVFTKR 483
: Db 468 SKFCCEFMKMLQGECTFLQAPAGVVDQGDHNEVSGSDVPASASIDISGYVVPCTA 527
: Qy 484 VVDKRYVSL-----GDILCVRL----- 502
: Db 528 LQPLQALDLPAEIVARAGRLTATVKSQVDGRIDCETLLDNKTPRTSFVDGAVLETNGP 587
: Qy 503 ----- 502
: Db 568 BRHNLSPDASQSTWMAAGPSSLTYAASAAGLEVRYVYAGLDHRAVPAPGVSPRSABEYTA 647
: Qy 503 ----- 502
: Db 648 PCSALYRFRBBAORLSLTGNFMFHBGLGPAPPSPGHVESANPFCGSSTLYTRTWS 707
: Qy 503 -----SPRO--VGF-----LPSRV-----PPA----- 517
: Db 708 VDAPSPAPQPDGFTSBSIPSRATPTPAAPLPAPADPSPTLSAPARGSPAGATARA 767
: Qy 518 ----- 517
: Db 768 PAITQTAHRRLFTYDGSKVPAAGLFESTCTMLVNVASVNDHRPGGLCHAPYQRYPA 827
: Qy 518 -----RVFHREBELVLRAGCTN---BRVPS 542
: Db 828 SFDAASFVRDGAAYTLTPRIIHAVAPDRLLENHPRLEBAAYRETCSRIGTAAYPLIG 887
: Qy 543 TPVVEBPQGFADLMH-----ATASLPEYNA-----TLQDALNDVQK---Lk 583
: Db 888 TGIYVPIGSPFDAERHNRPGDELYLELAAWFANRPCTPTLT--ITBDVATANTVIA 945
: Qy 584 ITLENAL---KITDGLTLPVRGLEMYR---GPSSGKTGTLIALBAAGKALVAVPTR 637
: Db 946 IELDSATVGRACAGCRVTP--GVQYQPTAGVPSGSRSTQK-----DVDVVVFTK 998
: Qy 638 ELREAMDRIRKPPSASATQHVAAAIL--RRAITBAGAPATVVIDECFMPPLVYVAIVAL 695
: Db 999 ELRNAMRRR---GFAAFPHTAARVTCGR-----VIDEAPSLP--PHLLILH-M 1043
: Qy 636 SPSSRIYVGVNHQGFIDPGQTSANMPLVDVQKCRRTFPCNKRCRPADVATTFPQS 755
: Db 1044 ORAAVHLLGDPNQIPALDFBAGI-VPAIRPDL--APTSWNHVTNRCPADV--CELJRG 1098
: Qy 756 LYPGCTTSGCAVASHVAPVYRNSQAOTLCFTQBEKSRHAGBAGMTYHBAQRTFASVI 815
: Db 1099 AYPMIQTTSRYLRSIFPMGEF---AVGQKLFYQAAKAA--PSSVTHEAGATYTTET 1152
: Qy 816 LHYNGSTAEQKLLAKS--HLVVGITRTHNLVYRDPGDIEROLNHSAKAEVFTDIPAP 873
: Db 1153 I---ATADAGLIGSSRAHAIVALTTRTEKCVIIDAG--ILREVGIS-DALVNNFPLAG 1207
: Qy 874 LEITTVKSEVQVR-----EVMATIPQSATPHGATILKKNPQDQDCGVAAAKTGY 928
: Db 1208 GEIGHQRPSVYIPRGPDANVDTLAAFPSCET--SAFHELEBELGHRP-APVAVALPCP 1264
: Qy 929 EVFGRAINVELAPDAPTPKPRAFQ--EGVQWVVTVAASNHQALQTLISYTRGSADL 987

```

```

Dh 1265 ELEGGGLYLPEGLTFCBSV-----TFELTDLVHCMAAPSQRKAVLSTLVERYGRRT--- 13117
Qy 988 PLHEAK-EDVKGMLNSLDRHMDVVTEDARDRAVFEIOLKFTQRCGTVEDLLEPD--DPY 10444
Dh 1318 KLVNASHSDVR---DSLARFLPALGPVQVTTCELYELEENAVEKEQODSDAVALDLCSDR 13744
Qy 1045 IRDIDFLMKYQOK-VSPKPIITGVGGGIIAASHKSLNFVLAAMTIRLIEE-LTRGSRVR 11020
Dh 1375 VSRITTFQKCNKKTGTGTLIHGKVGQGISMSKTFCALPGPWFAIKALIALLPQGVF 14344
Qy 1103 YSNGLPDEE-EAMLEAKINOVPHATFVSADMTBERDTAHNNTSEILFALERIGTPAAA 11614
Dh 1435 YGDAFDDTVFSAVAANAASV-----FENFSEBDSIQNNFSLGLECAIMEBCMPWL 14894
Qy 1162 VNLFRERCGKTLPAKIGSVVEVDGL-----LDGSAANTPCRNITFSAAVMLTL--FRGV 12144
Dh 1490 IRLY-----HLIRSAMILQAPKESLRGFMKKHSGEPGTLTMTVTWMAVITHCYDFRDL 15434
Qy 1215 KFAFPKGDSDS-LGSHYLRDASRLHNGERYKTHLVEVQKIVPYIGLLVSAEQVVL- 12727
Dh 1544 QVAAPFKGDSTVLCEYRQSPGAVALIAGGILK--LKVDPRPILGYAGVVAPGLALP 16000
Qy 1273 DPVRS 1278
Dh 1601 DVARFA 1606

RESULT 6
US-08-471-971-1
: Sequence 1, Application US/08471971
: Patent No. 6287759
: GENERAL INFORMATION:
: APPLICANT: Tsarev, Sergei. A., Emerson,
: APPLICANT: Suzanne U., Purcell, Robert H.
: TITLE OF INVENTION: Recombinant Proteins Of
: TITLE OF INVENTION: A Pakistani Strain Of Hepatitis B And Their
: TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
: NUMBER OF SEQUENCES: 107
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/471,971
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US08/316,765
: FILING DATE: 03-OCT-1994
: CLASSIFICATION: 435
: FILING DATE: 18-SEP-1992
: APPLICATION NUMBER: US07/947,263
: PRIOR APPLICATION DATA:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Richard W. Bork
: REGISTRATION NUMBER: 36,459
: REFERENCE/DOCKET NUMBER: 2026-4032US2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 758-4800
: TELEFAX: (212) 751-6849
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1693 AMINO ACID RESIDUES
: TYPE: AMINO ACID
:

```

;	STRANDEDNESS: UNKNOWN
;	TOPOLOGY: UNKNOWN
US-08-471-971-1	
Query Match	5.7%; Score 500; DB 3; Length 1693;
Best Local Similarity	21.2%; Pred. No.1.9e-34;
Matches 371; Conservative 186; Mismatches 575; Indels 614; Gaps 72.	
QY	2 YAKATDVARVYAADVAAYANTVLOQRAVKLDFAPPLKALETLHRLYY- -LRFGGTLRPT 59
DB	6 FIKAPGITTALIEQALAAANSALANAVVREFPLSHQOIEILINIMQRLQVREBFV--W 63
QY	60 QHPLIAGHORAEVNLNPARGRS--TYLEIGPSLHSLAKHGAPNAPADVHGC--TKYGT 117
DB	64 NHPI-----QRYINHELELYCARAGRCIEIGAPRS--INDPNV-----VHRCPLRAG 112
QY	118 RDGSRHITALE-----SRSAVGRPEPFKADASILANGIASRTFCVDVGSCAFKSVVG 170
DB	113 RDVGRWTTAPPRGPANCRBSALAGLP-----AADRYTCFDSGSGCNFPATG 160
QY	171 IANHSLYDVTLEBLANAFENHGLMVAFAFMMPBELLYMDNVNVAELGYRPHVIEBPMAV 230
DB	161 IALYSLHDMSPSDVABAMFRRGMTRLYALALHPREVL-----PEPTY 203
QY	231 KDCAPQGGDLRLHPELDFINESGBERRIERLAAGSISRRAVVISGDDMDGDAYLD-- 287
DB	204 RTAAY-----LLIH-----DGRV-----VVTYEGDTSAG--YMHDSVN 235
QY	288 FHTWALVLYVANYPTPPGFSLHIEVORRHGSSIELRTAP-----PGDRMLAVYPR 339
DB	236 LRSM-----ITTKVTDGDPVILYRVAIGCFVULLTAABPSPMPVYPRSTEVYR 290
QY	340 TSGGLCRIPNIFYAYDASGTEHKILTISOHKVNNLNFMOQTRPEKELVDMTVLMSAPAR 399
DB	291 SIFPGGPRSLF--PTSCSTKSTHVAVDHIMDLMLFGATLDQAPC--CSRLMTYLAGI 347
QY	400 LRAIVASSEVTESSWNISPADLVATVVSIVY-----LHIERRAAVAV 443
DB	348 SYKTVGTLVANBGMNASBDALTIVITAAVYLTICHOYLRTQAIASKGMRRLEREHAKFI 407
QY	444 -----KXAD-----DVGRBSPFWSL-----GV 463
DB	408 TRLYSMLEFBSGRDIYIFGRQLEFYAQCRMLSAGFHLDPVLVDEBSAPCHCKTAIRAV 467
QY	464 LGSCC-----GLNLKGTDVVFYKR 483
DB	468 SKFCFPMKMLQECTCFLOPARBGVVGQGHNEAVESGDVDPAESISDISGSYVPEGA 527
QY	484 VVDKYRVHSL-----GDIIICDVL----- 502
DB	528 LQPLYQALDIPAEIVABAGRLTAVKVSQVNDGRIDCEFTLGNKTFRTSFVDGVALVETGP 587
QY	503 ----- 502
DB	588 ERHNLSPDASGSTMAGPFSLTYAASAAGLEVRVYAAGLDHRAVAPGVSPBSAAGEVTA 647
QY	503 ----- 502
DB	648 FCSALYRPNRAQGLSLTGNFWPPEBGLGFPAPSPGHVMSANPFGBSTLYTKRTWSE 707
QY	503 -----SPEQ--VGF-----LPSRV-----PPA----- 517
DB	708 VDAVPSPAQPOLGFTSBSIPSRAATPTPAAPLPPAPADPPTLSAPARGBPAPQATARA 767
QY	518 ----- 517
DB	768 PAITHQTAHRRLFTYPDGSKVAGSLFESTCTWLNVASVNDHRPQGLCHAFQRYPA 827
QY	518 -----RVFDRLELEVLREAGCYN-----ERPVS 542
DB	828 SFDAASPYMRDGAALYTLTPRPILIHAVAPDRLHNPKRLERAAARETCSRLGTAYAPLLG 887
QY	543 TPVPEBQGFADADIMH-----ATAASLEPYRA-----TLQAGLNTDVQK--LK 583

```

Db      | 888 | TGIYVPIGSPFDAMERNRPGDELYLPBLAARWFEANRPTCEPTLL--TTEVVARPANLA 945
Qy      | 584 | ILENAL--KTIDGLTSPVRLGLEMYE---GPPSGKGTGLTIALAALAGKALYVAPTR 637
Db      | 946 | IELDATDVGRACACGRVTP--GVVOYOTAGVPGSGKRSITQA-----DVDVVVPR 998
Qy      | 638 | EUREAMDRRIKPPSASATQVALAIL--RRATAEGAPFATVVIIDECFMPPLVVAIVHAL 695
Db      | 999 | ELRNAMRRR---GFAAFPTHTAARVTOGRH-----VVIDEAPSLP--PHLLILH-M 1043
Qy      | 696 | SSSSRIVLVGDVHOGFIDFOGTSANMPLVRODVVOCRRRTNOCFCGADVVATTFPS 755
Db      | 1044 | ORAAVHLIDPNQIPALIDFEIAGL--VPAIRPDL--APTSMMVHTRCADPA--CELLIG 1098
Qy      | 756 | LYPGCTTSGCVASISHVAPDYRNSQOTLCFQOEKSRHAGAMTVEAOGRTFASVI 815
Db      | 1099 | AYPMTQTSRVLRSLFWGEP---ANGQKLVFTQAANKAN--PGSVTVEAOGATYTETT 1152
Qy      | 816 | LHYNGSTAEQKLLAEKS--HLLVGITRATNHLIYINDPTGDIERQLNHSKAEVFTDIPAP 873
Db      | 1153 | I---ATADARGIIGSSRAHAIVALTREHTEKCVIIDAPG--LIREVGIS--DAIVNFFLAG 1207
Qy      | 874 | LEITTVKPESEVQNR-----EWMAITPPQSATPHGAIHILRRKRFQDPOCCVAAKATGY 928
Db      | 1208 | GEIGHORPSVIRPGNDANVDTLAAPPSCET--SAFHBLABELGHRP--APVAAVLPCCP 1264
Qy      | 929 | EVFGRAKINVELAEBDAPTRPKPHAFQ--EGVQWVKVTNASKHQAOLQTLRSYTKRSADL 987
Db      | 1265 | ELEQGLIYLPOLITTCDSVY---TRELDIYHCRMAASQKRAVSLTVGKRGRT--- 1317
Qy      | 988 | PLHEAK-EDVKMELNSLDHRHMTVTEDARDRAVFTQKLTORGQTVEDLLEPD--DPY 1044
Db      | 1318 | KLYNASHSDVR---DSLAFRIPAIGVQVTCCELYELBEMAVEKGDGSAVELDLCSDH 1374
Qy      | 1045 | INDIDPLMTQOK-VSPKPIINGKVGOGIAHNSKSNFLPAMIRLLEB--LIRTSRTYR 1102
Db      | 1375 | VRITTFQKDCNKFPTTGETIAGKVGOGISAMSKTICALGPPFRALIEKAILLALPQGVF 1434
Qy      | 1103 | YANGLPDEE-EMALTEAKINQVPHATPVADWTEFTJANNNTSELLFALLERIGTPAA 1161
Db      | 1435 | YGDAPDPTVPSAAVAAKASMT-----FENDPSEFTSTONNBSLGIECAIMECGRPOL 1489
Qy      | 1162 | VNLFRERCCKRTLRAKGLSVEVDGL-----LDGGAAMPCCRNITFSAAVMLTL--FRGV 1214
Db      | 1490 | IRLY-----HLIRSAMILQAPKESLRGFWKHSGBPTILMTVMNMAVITHCYDFDL 1543
Qy      | 1215 | KRAAFKGDSTL-CGSHYLRFDASRLMGERYKTKHLYEVQKIYVYIGLTVSABGVV-- 1272
Db      | 1544 | QVAAPFGDSDIVLCSEYRQSPGAALVLAGGGLK--LKVDPRITGLVAGVVAVPLGALP 1600
Qy      | 1273 | DVPRSA 1278
Db      | 1601 | DVPRA 1606

```

RESULT 7
 US-09-402-776-1
 Sequence 1, Application US/09402776
 Patent No. 6458562

GENERAL INFORMATION:
 APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
 APPLICANT: Tearev, Sergei. A., and Robinson, Robin A.
 TITLE OF INVENTION: Recombinant Proteins Of
 TITLE OF INVENTION: A Pakistani Strain Of Hepatitis B And Their
 TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
 NUMBER OF SEQUENCES: 111
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: MORGAN & PINNEGAN
 STREET: 345 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: USA

```

Qy      | 2 | YAKATDVARVVAADVAAYVNLQORAVYXLDPAAPKALETTLRLYYP--LAPKGGTLPT 59
Db      | 6 | FIKARGITTAIRQALAAANSAANAVVVRPLSHQOEILINTMQPQLVREVF--W 63
Qy      | 60 | OHPILAGHORVAEVLNPARGRS--TVLEIGSLHSALKLHAAPNAPVADYHGC--TKYGT 117
Db      | 64 | NHPI---QRYVHNELEBYCRASGRCLGIGHPRS---INDNPV---VHRCFLRPA 112
Qy      | 118 | RDSRHITALE-----SRVATGRPEPKADASLLANGIARTPCVQVSGCAPKSRVG 170
Db      | 113 | RDVQWHTYAPTRGPANCRRLRGLP-----ADRTTCFDFSGCNPALGTG 160
Qy      | 171 | IANSLYDVTLEELANAFENHGLHNVRAFMNPRBELLYMDNVNVELGYRPFVIEPMAY 230
Db      | 161 | IALYSLHDSPSDVAEAMFRHGMTRLYAALHLPVYTL-----PGTY 203
Qy      | 231 | KOCARQGGDLRIHFFELDPINESQRRIRBLAARGSYSRRAYIFSGDDDMGDAYLHD--- 287
Db      | 204 | RTASY---LILH-----DGRV-----VVTYSGDTBAG--YNNHDSN 235
Qy      | 288 | FHTWLAVLLVRYVPTPFGSLHIEVQRHGSISBLRTRAP-----PGDRMLAVVR 339
Db      | 236 | LRSW-----IKTKYTGDPHLVIERVRAIGCFULLITAAPEPSMPYVTFYRSTEVYR 290
Qy      | 340 | TSOGLCRIPNIFYVADAGTEKHTILTSQHKVNLMLNFQTRPEKELYDMTVLMSPARAR 399
Db      | 291 | SIFPGGTRSLF--PTSCSTKSTFAVPAHINDRLMLFATLDDQAF-C-SRLMTYLRGI 347
Qy      | 400 | LPAIVVASEVTSSNNISPADLVRTVLSLYV-----LHTRRRRAAVAV 443
Db      | 444 | -----KTAQD-----DVFGSTPWSL-----KHV 463
Qy      | 408 | TLYSMLPEKSGRDYIPGRQLEFYACRRMLSGFHLDPVUVPPBSAPCHCRITARKAV 467
Db      | 464 | LQSCC-----GLRNLKGTDVVFTYR 483
Qy      | 468 | SKFCFEMKVLQGECTFLQPAEGVVGDOCHDNEAVGSDVDPBASALISDISYVVPGR 527
Db      | 484 | VVDKRYVHSL-----GDIIICVRL----- 502

```

ZIP: 10154
 COMPUTER READABLE FORM:
 MEDIUM TYPE: FLOPPY DISK
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: WORDPERECT 5.1
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/402,776
 FILING DATE:
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/840,316
 FILING DATE: 11-APR-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Richard W. Bork
 REGISTRATION NUMBER: 36,459
 REFERENCE/DOCKET NUMBER: 2026-4255
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 758-4800
 TELEFAX: (212) 751-6849
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1693 AMINO ACID RESIDUES
 TYPE: AMINO ACID
 STRANDEDNESS: UNKNOWN
 TOPOLOGY: UNKNOWN

US-09-402-776-1

Query Match 5.7%; Score 500; DB 4; Length 1693;

Best Local Similarity 21.2%; Pred. No. 1,96-34; Matches 371; Conservative 186; Mismatches 575; Indels 614; Gaps 72;

D	b	528	LQPLYQALDLP	PAEIVADAGRLTATVAVS	QYVDGRID	CE	TLGNKTPRTS	PVDGAVLE	TNGP	587
Q	y	503	-----	-----	-----	-----	-----	-----	-----	502
D	b	588	ERNHLS	PDASQSTMAAGP	FSLT	TYAASAA	GLEVR	VVAAGLDHRA	VPFAGVSPRSAPGE	647
Q	y	503	-----	-----	-----	-----	-----	-----	-----	502
D	b	648	FCSALY	RFNBEAQRSL	TGNWFPE	CGILGP	APFSPG	HVWESANP	CGSESTLYRT	707
Q	y	503	-----	SREQ	-----	VGF	-----	LPSRY	-----	517
D	b	708	VDAP	SPAQDPLG	FTSEPS	IPSRAT	PTPAAP	LPDPAP	DPSPFL	517
Q	y	518	-----	-----	-----	-----	-----	-----	-----	517
D	b	768	PAITHQ	TARRRRL	FTY	PDGSK	VAVAGSL	FESTCT	WLVN	827
Q	y	518	-----	-----	-----	-----	-----	-----	-----	542
D	b	828	SFDA	SPVMBDGA	AAAYTL	TPRP	II	HAAV	PDYRL	887
Q	y	543	TPPVEE	QGFADLMH	-----	ATA	SLPE	RYA	-----	583
D	b	888	TGIYQ	VEIPGSP	DFMERN	HRBG	DELY	PELA	AFMEAN	945
Q	y	584	ITLENAL	-----	KTIDGL	TSVPR	GLEM	Y	-----	637
D	b	946	IELDS	ATDIDVGR	ACAC	CRVTP	-----	GVVQ	IFTG	998
Q	y	638	ELRE	AMDRRIK	PSASAT	QVHAL	IL	-----	RRAT	695
D	b	999	ELRNA	MRR	-----	GFAA	FTPH	TAAR	VTQGR	1043
Q	y	696	SPSS	RIYLVG	VQHG	IF	ID	FOG	TSA	755
D	b	1044	QRA	TVHLG	LPNOI	PAID	FE	HAGL	-----	1098
Q	y	756	LYPC	CTTSGC	VA	SHVAP	DYR	NSQA	OTL	815
D	b	1099	AYPI	QITSR	VLRS	FWGEP	-----	AVG	KLV	1152
Q	y	816	LHY	NGSTAE	OKLAE	KS	-----	HL	VG	873
D	b	1153	II	-----	ATD	ARG	L	QSSRA	HAIV	1207
Q	y	874	LE	ITTVK	PESE	VO	RN	-----	EW	928
D	b	1208	GEI	GHQ	PSV	IPRG	ND	AV	DTL	1264
Q	y	929	EV	FGRA	KIN	TEL	EPD	AT	PKP	987
D	b	1265	ELE	GL	LYL	PO	EL	TT	CS	1317
Q	y	988	PL	HEAK	-----	EDV	GR	ML	NS	1044
D	b	1318	KL	VNASH	SDV	-----	DSL	AR	F	1374
Q	y	1045	IR	ID	FL	MKT	QOK	-----	VS	1102
D	b	1375	VS	RT	TF	FO	KDN	KFT	Y	1434
Q	y	1103	Y	SN	GL	PD	PE	-----	BA	1166
D	b	1435	Y	GD	AF	DD	TV	FS	AA	1489
Q	y	1162	V	N	L	F	E	R	C	1214
D	b	1490	I	R	L	V	-----	HL	S	1543
Q	y	1215	K	P	A	E	K	G	D	1272

```

Db      1544  QVAAPKDDSDIVLCESEYRSGAAGVAVLJAGSLK---LKVDPRPIGLVAGVAVAPQLGALP 1600
QY      1273  DPVPSA 1278
          |||
Db      1601  DVVRFA 1606

RESULT 8
PCT-US93-08849A-1
/ Sequence 1, Application PC/TUS9308849A
/ GENERAL INFORMATION:
/ APPLICANT:
/ TITLE OF INVENTION: Recombinant Proteins Of
/ TITLE OF INVENTION: A Pakistani Strain Of Hepatitis B And Their
/ TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
/ NUMBER OF SEQUENCES: 98
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSER: MORGAN A FINNEGAN
/ STREET: 345 PARK AVENUE
/ CITY: NEW YORK
/ STATE: NEW YORK
/ COUNTRY: USA
/ ZIP: 10154
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: FLOPPY DISK
/ COMPUTER: IBM PC COMPATIBLE
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: WORDPERFECT 5.1
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: PCT/US93/08849A
/ FILING DATE: 17-SEP-1993
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US07/947,263
/ FILING DATE: 18-SEP-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: William S. Feiler
/ REGISTRATION NUMBER: 26,728
/ REFERENCE/DOCKET NUMBER: 2026-4032 PCT
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (212) 758-4800
/ TELEFAX: (212) 751-6849
/ INFORMATION FOR SEQ ID NO: 1:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 1693 AMINO ACID RESIDUES
/ TYPE: AMINO ACID
/ STRANDEDNESS: UNKNOWN
/ TOPOLOGY: UNKNOWN
/ PCT-US93-08849A-1

Query Match      5.7%; Score 500; DB 5; Length 1693;
Beet Local Similarity 21.2%; Pred. No. 1.9e-34;
Matches 371; Conservative 186; Mismatches 575; Indels 614; Gaps 72

2 YAKATDVARVYAADVAAYANVLOQRAVVLDPAPPLKALETLRLTYLP--LRFKGGTLPT 59
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
6 PIKARGITTAIEQALAAANSAANAAVAVVRPLSHQGIILLINMQPQLVRRPEVF--W 63

60 QHPILAGQRAVEEVLTHNPAGRS-TVLEIGPSLHSAALKGAPAPVADYGC--TKYGT 117
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 NHPI---QRYIHNELELYCRARSGRCLIEIGAPRS--INDPNV---VHRCPLRPAG 112

118 RDGSHITALE-----SRSVATGRPEFKADASILANGIASRTFCVGVGSCAFKSRVG 170
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
113 RDVGRWYTAIPRGPAAACGRSALRLP-----ADRTYCFDGFSGCNPAETG 160

161 IALYSIHMSPSDVDEAFMRHGMRLYLAHLPEVLL-----PGTY 203

231 KDCARQGGDLHLHPELDFINESQGRIRLELAARQSYSRRAVIFSGDDDWGAYLHD--- 287
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
204 KTAAY---LTLH-----DGRV-----VVTYGGDTAG--YNHDSN 235

```


Query Match 5.7%; Score 500; DB 5; Length 1693;
 Best Local Similarity 21.2%; Pred. No. 1.9e-34;
 Matches 371; Conservative 186; Mismatches 575; Indels 614; Gaps 72;

2 YAKATDVAAYAAADVAAYANVLOQRAVXLDPAAPLKALETLRLYYP--LRFKGGTLPT 59
 6 FIKAPGITTALEQALAAANSALANAVVRPELISQOIEILINMQPQVLRPREVF--W 63
 60 QHPIAGQRYAAEVLNHPARGRS--TVEIGPSLHSLKLGAPNAPADYHGC--TKYGT 117
 64 NHPI---QRYIHNELELYCRASGRCLGHPRS---INDNPV---VHRCFLRAG 112
 118 RDGRSHITALF-----SRSAVATGRPEFKADASILANGIASRTPCVDSVSCAFKSVG 170
 113 RDVGRWYTAFTPRGPANCRSALRGLP-----AADRTYCFDGFSGCNFPAETG 160
 171 IANHSYDVTLEELANAFENHGLHWVRAFMHMPBELLYMDNVNVALCYRPHVIEBPNAV 230
 161 IALVSLHMSPSDVAEAMFRHGMTRLYAALHLPEVLL-----PQCTY 203
 231 KOCAPQGGDLRHFPELDFINESQERRIERLAAGSYSRRAVIFSGDDDKGDVYLD-- 287
 204 RTASY---LTIH-----DGRVY-----VITYEGDTSAG--YMHDSN 235
 288 FHTMLAYLLVRYNPTPPGFSLHIEVQRHSSIELRTRAP-----PGDRMLAVYPR 339
 236 LRSM---IRTKYTGDPHPLVIERVRAIGHFVLLTLAABPSPMPYVPRSTEVYR 290
 340 TSQGLCRIPNIYYADAGSTEKHTILTQHKVNMILNFMQTRPEKELVDMTMSFAPAR 399
 291 SIFPGGTSLF--PTSCSTKSTPHAVPAHIMDRMLGATLDDQAF--CSRLMYLYRG 347
 400 LRAIVASEVTSSWNISPADLVRTVSLYV-----LHIERERRAAVAV 443
 348 SYKTVGLVLANEGNNASEDALTAVITAAVYLTICHQRYLTOALSCKGRRLEREAQRFI 407
 444 ---KTKAD-----DVGSTSPWESL-----GV 463
 408 TRLSWLFESKGRDITPGRLQLEFYAQCRMLSAGFHLDPRLVDFESAPCHCRTAIRAV 467
 464 LGSCC-----GLANLKTGVVFTKR 483
 468 SKPCCFMKMLGOECTCFLOPAEGVVGDOGHNEAEGSDVDPAESAIDIGSYVVPOTA 527
 484 VVDKRYVSL-----GDIIICDVL----- 502
 528 LQPLYQALDLPAEIVARAGRLTATVKSQVODRIDCETLLGNKTRTSFVDAVLETGP 587
 503 --- 502
 588 ERHNLSPDASQSTMAAGPFLTYAASAGLEVRYVAAGLDHRVAPGVSPRASGEVTA 647
 503 --- 502
 648 FCSALYRNRRAQRLSTGNFWFHEBGLGPPAPSPGHVESANPCGGBSTLYRTTWSE 707
 503 ---SPBQ---VGF---LPSRV---PPA----- 517
 708 VDAVPSPAQPOLGFTSEPSISRAATTPPAAPLPPPADPPTLSAPARGBPARGATARA 767
 518 --- 517
 768 PAITHQTAHRRLLETPYDGSKVPAAGLFEESTCTWLNVASNVDRPGGLCHAFQRYPA 827
 518 ---RVFDRSELEVLRAGCVN---ERPVS 542
 828 SPDAASFMRDGAAYITLTPRPIIHAVAPDYRLBNPKLEAARRETSSRLGTAYPLIG 887
 543 TPVEEPOGFDADLWH---ATAASLPEYRA---TLQAGINTDVXQ---LK 583
 888 TGIQVPIGSPFDAMERNHRPQDELXLPBELAARWFANRPTCPTLT--ITEDVATANIJA 945
 584 ITLENAL---KTIDGLTISPVRGLEMYE---GPPSGKTGTLIALLEAAGKALYVAPTR 637

DB 946 TELSDATDVGACAGCRVTP--GVVQYQFTRAGVSGSKSRITQD-----DVDVVVUTR 998
 QY 638 ELREAMDRIKPPASATQYVALAIL--RRATAGAPPATVVIDECFMPPLVYVAIVHAL 695
 DB 999 ELRNAMRR---GFAAFPTHAAVTOGR-----VIDABSLP--PHILLAH-M 1043
 QY 696 SPSSRIVLVGVNHOIGFIDPGSTANMPVYADVVKQCRRRFPNOTKRCBPADVATTFFQS 755
 DB 1044 QRAATVHLGDPNPIPAIDFENAGL--VPALRPDI--APTSWHTYHRCPADV--CELIRG 1098
 QY 756 LYPCCTTSGCVASISHVAPDYRNSQAOTLCFTQEBKSRHAGBAGMTYBAQGTFAV 815
 DB 1099 AYPMIOITTSRYRLBFLMCEP---AVQKLVFTQAAPAN--PSVYTHEQAGATYTTT 1152
 QY 816 LHYNGSTAEOQLAELS--HLVGIITRTHNLVYRDPTGDIROJLNSAKAEVTDIPAP 873
 DB 1153 IIT--ATDANAGLIQSSRAHIVALTRHTKCVIIDAAG--LLREVGIS--DAIVANNFLAG 1207
 QY 874 LEITTVKPSSEVQNN-----EWAATTIPQSATPHGAIHLKRNFGDOPDCGVALAKTY 928
 DB 1208 GEIGHQRPSTVIRGNPDANVDTLLAFPPSCBI--SAFHELABIDGHRP--APVAALVLPCCP 1264
 QY 929 EVFGGRKINVELAPDTPKPHRAFO--EGVQWYKVTNASKHQALQTLISRYTKRSADL 987
 DB 1265 ELBOGLYLPQELTTCOSV---TFELTDIVHCRMAAPSORKAVLSTLVGRYGRRT--- 1317
 QY 988 PLHBAK--EDVCRMNLSDRHMWVTEDARDRAVEFTQKFTORGCTVEEDLEBD--DPY 1044
 DB 1318 KLVNASHSDVR---DSLARFIPATGPVQVTTCELYLEBAAVEKGQDSANVLEBDLCSRD 1374
 QY 1045 IRDIDFLAKTQOK--VSPKPIITGVGOGIAHSGSLNFTLAWIRILEE--ILRTGSRFVR 1102
 DB 1375 VSRITFPQKCNKKTETBETIAHGVGOGISAMSKTFCALFGPMFRAIEKAILALPQGVF 1434
 QY 1103 YSNGLPBEE--EAMLEAKINOVPAITVSAADMTPEPDAHNTSELFPALIERIGTPAA 1161
 DB 1435 YGDAPFDITVPSAAVAAKASNV-----FENDFSEDTQNNPISGLBCALNEBCMPWL 1489
 QY 1162 VNLFRERQKRTLRAGKGVSEVVDL-----LDSGAAMTPCRRNTIFSAAVMLTL--FRGV 1214
 DB 1490 IRLV-----HLHSAMILQAPKESLNGFMWKGSGBPOTLLMNTVMNAVITHCHDFPDL 1543
 QY 1215 KPAAPKGGDSL--LGSHTLRPDASRLMNGERYKTKHLKVEYOKVVPYIGLLVSAEQVVL- 1272
 DB 1544 QVAAFKGDDSVLCSBYROSFGAAVLLAGCLK---LKVDPRPIGLVAGVAVPAGLALP 1600
 QY 1273 DPVRSA 1278
 DB 1601 DVVRPA 1606

RESULT 10
 US-08-478-507-7
 ; Sequence 7, Application US/08478507
 ; Patent No. 6120988
 ; GENERAL INFORMATION:
 ; APPLICANT: Reyes, Gregory R
 ; APPLICANT: Yarbough, Patrice O
 ; APPLICANT: Bradley, Daniel W
 ; APPLICANT: Krawczynski, Krzysztof Z
 ; APPLICANT: Tam, Albert
 ; APPLICANT: Fry, Kirk B
 ; TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
 ; NUMBER OF SEQUENCES: 20
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: DeHlinger & Associates
 ; STREET: 350 Cambridge Avenue, Suite 250
 ; CITY: Palo Alto
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94306

```

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/279,823
FILING DATE: 25-JUL-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholez, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ. ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-507-7

Query Match      5.4%; Score 479; DB 3; Length 1693;
Best Local Similarity 21.2%; Pred. No. 1.4e-32;
Matches 369; Conservative 188; Mismatches 577; Indels 610; Gaps 73;

```

```

291 61FGPGCTPSLF--PTSCSTKSTPAVPAHMDRLMFGATLDDAFC--CENLMTYLLKCI 347
400 LRAIVASBRYTESSNNISPADLVRTVSLVY-----LHIIERRRAAVAV 443
348 SYKTVVGLTVANBGNASBDLTAIVTAAYLTIHQRYLRTQALSKGRRLERBAHQFI 407
444 -----KRAKDVPG-ETSPWESLHVLS----- 466
408 TRLYSMLPEKSGRDYIPGRQLEFYAQCRWLSAGFHLDPRLVFDDESAPCHRTAIRKL 467
467 -----CC-----GLNLTGDTVVFTR 483
468 SKPCCPKMKLGECTCFIQAPEGAVDGDHNEAVEGSDVDPAESAIDISSTVYVPGTA 527
484 VVDKRYVHSL-----GDIIICDVL-----SP 504
528 LQPLVQALDLPALVIAVAGRLTATVKGQVDRICETLLGNKTRFRTSPVDAVLETNGP 587
505 EQ-----VGFPSRPV----- 515
588 EKHNLSPDASGTMAAGPSLTLYAASAAGLEVRVYAAGLDHRAVEAPGVSPSAGEVTA 647
516 -----PARVHD----- 522
648 FCSALYRVRREAQRSLGNLMFHEBGLGLPAPSPGHVBSANPCGCESTLYTRTWSE 707
523 -----REBLEVLREAG-----CYNERVPSTPVYBEPG----- 551
708 VDAVSPARPDLGFWSEPSIPSRATPTLAAPLPPADDPSPSPAPLABASGATGA 767
552 -----PADJMHAT----- 560
768 PAITHQTAHRRLLFTPDGSKVFAGSLFESTCTVLVNASVNDHRPGGLCHAFQRYPA 827
561 ---AASL-----PEYR-----ATIQG-----LN 576
828 SPDAASFWNRDGAAYTLTPRTIHAIVADYILHNPKRLBAAYETGSRGLTAAYPLIG 887
577 TDVKQ----- 585
888 TGIVQVPIGSPDABERNRHPGDELYLPELAWMFEANRPTRPTLTITEDVARTALAI 947
586 LENAL---KTIGGLTSPVKGDMTE---GPGSGKTGTLTALBAAGKALYVAPTRBL 639
948 LBSATDVGCAGACRVTG--GVVOYQTAGVSGSGSRITGA-----DVDVVVPTREL 1000
640 REAMDRRIKPPASATQHVLAAIL--RRATAGAPFATVVIDCFMFLPVVAIVHALSP 697
1001 KNAWRRR---GFAAFTPTTAAYTQGR-----VIDEASRLP--PHILLH-MQR 1045
698 SSRIVLVGDVHIGFIDFGSTANMPLVVDVVKQCRRTFNQTKRCPADVVAATTFPQSLY 757
1046 AATVHLGDPNQIPALDFEHAGL-VPAIRPDGLP--TSMWHTYHRPADV--CELRGAY 1100
758 PGCTTSGCVASISHAAPDYRNSOAGTLCTPQEBKSRHAGAGMATHSOGRTFASVILH 817
1101 PMIQTSRVLRLFWGEP---AVGQKLVFTAAAPAN--PSSVYHKGAGATYETTTII 1154
818 YNSTAOKLBAKS--HLVGTIRHTNLYLRDPGDIROLNSAKALEVTDIPAPLE 875
1155 ---ATDANGLQSSAAHAIVALTRHTEKCVIIDABG-LIREVGIS-DALVNNPFLAGE 1209
876 ITTVKPSBEVQRN---EVMATTPQASATPHGAILHLKRNFGDDCCGVALAKTYEV 930
1210 IGHQPSVLPKGNPDANVDVTLAAPPSCQI--SAFQQLABRIGHRP--VVAVALPCCPBL 1266
931 FCGRAKINTELAPDAPTRKPHAFQ--EGVQWYKVTNASKHQAOLQTLISRYTKRSADLU 989
1267 EOGLLYLPQELTTCQSVV---TFBLTDTIVHCRMAAPSORKAVLSTLVRIGGRT--KL 1319
990 HEAK--EDVGRMLNSLDHRMDVTVEDADRAVAFETOLKFTQSGTVEEDLEPD--DPYIR 1046

```



```

Db      1320  YNASHSVR---DSLARFIPAIGPQVYTTCELVELVEMAVEKGDDGSAVLELDCNRDVS 1376

QY      1047  DIDLMKTQOK-VSPKPIINTKXGOGGIAHKSILNFVLAAMIRLEE-ILFTGSRTRYRS 1104
Db      1377  RITTFQXCKNFTTGTETIAHGRVGGISAWSEKTCALCPMPFRAIEKXILLLPQGVYFG 1436

QY      1105  NGLPDEE-EAMLEAKINQVPHATFVSADMTVEPDTAHNNTSELPAALLERIGTPAAVN 1163
Db      1437  DAFDDTVAFAVAAKASGMV-----FENDFSEFDSTQNNFSLGECAIMECGMPWLIR 1491

QY      1164  LFRERCGRITLRAGLGSVEVDGL-----LDGSAAMTFPCRTTITSAAMMLTL--FRGYKF 1216
Db      1492  LY-----HLIRSAMILLQAPKESIRLGFPMKKHSGEBGTLTLMATVNMMAVITHCYDFRPFQV 1545

QY      1217  AAFKGGDSL-LCGSHYLRFDAISRLLMGERRYTKHLKLVQOKIVPIYIGLVAESAEOVL--DP 1274
Db      1546  AAFKGGDSBIVLCSEYRQSGPAAVILAGGLK---LKVDFFRPIGLYAGVVVAPGIALPDV 1602

QY      1275  VRSA 1278
Db      1603  VRFA 1606

```

RESULT 11
 US-09-128-275A-7
 Sequence 7, Application US/09128275A
 Patent No. 6229005
 GENERAL INFORMATION:
 APPLICANT: Reyes, Gregory R
 APPLICANT: Yarbough, Patricia O
 APPLICANT: Bradley, Daniel W
 APPLICANT: Krawczynski, Krzysztof Z
 APPLICANT: Tam, Albert
 APPLICANT: Fry, Kirk E
 TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
 TITLE OF INVENTION: No. 6229005-A/No. 6229005-B Hepatitis Viral Agent
 NUMBER OF SEQUENCES: 20
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Dehlinger & Associates
 STREET: 350 Cambridge Avenue, Suite 250
 CITY: Palo Alto
 STATE: CA
 COUNTRY: USA
 ZIP: 94306
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/128,275A
 FILING DATE: 03-AUG-1998
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/279,823
 FILING DATE: 25-JUL-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/681,078
 FILING DATE: 05-APR-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/505,888
 FILING DATE: 05-APR-1990
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/420,921
 FILING DATE: 13-OCT-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/367,486
 FILING DATE: 16-JUN-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/336,672
 FILING DATE: 11-APR-1989
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/208,997
 FILING DATE: 17-JUN-1988

[illegible]

```

Db 708 VDAVSSRRPDLGFMSEPSISBRATPTLALPLPPAPDPSPSPAPALABASGATAGA 767
Qy 552 -----FDADIWHA----- 560
Db 768 PAITHQTAHRRLFTYPPDGSKVFAGSLFESTCTVLNANVNDHRPGGLCHAFQRYPA 827
Qy 561 ---AASL-----PEVR-----ATLQAG-----LN 576
Db 828 SPDAASFVWRDGAAYTLTPRIHAIVADYRLHNPKRLEAAYETCSRLGTAAYPLL 887
Qy 577 TDVQK-----LKIT 585
Db 888 TGIYQVPIGSPSPDAMERHNRDELYLPDLAARMEANRPRTPLITEDVARTANLAI 947
Qy 586 LENAL---KITDGLTSLSPVRLMEYE---GPPGSKTGTLAALBAAGKALVAPTEBL 639
Db 948 LDSATDVGRACAGCVTP--GVVQYQFTAGVPGSGKSRSITQA-----DVDVVVPTREL 1000
Qy 640 REAMDRRIKPPSASTOHVALAIL--RRAATAGAFATVYVDECFMFLVYVAIVHALSP 697
Db 1001 RNAMRR---GFAATPPTAAVTOGR---VVIDEADSLP-PHLLILH-MQR 1045
Qy 698 SSRIVLGVDVHQIGFIDFGTSANMPLVVDVVKOCRRRTFNQTKCPADVATTFFOSLY 757
Db 1046 AATVHLDDPNQIPAIDEFHAGL-VPAIRPDLGP--TSWMHTHMPADV--CELIRGAY 1100
Qy 758 PGCTTSGCVASISHVADYRNSAQOTLCFTOEBSRGAGBAMTVHEAOGTFASVLIH 817
Db 1101 PMIQTTSRVLRSLFNGEP---AVGQKLVFTQAAPRAN--PGSVTVHEAOGATVETETI 1154
Qy 818 YNGSTAEOKLAEKS--HLVGTIRHTNHLRYRDTGDIEROLNMSAKAEVTTDIPAPLE 875
Db 1155 ---ATADAGLIGSSRAHAIVALTRHTEKCVIIDAPG-LRREVGIS-DAYVNFPLAGE 1209
Qy 876 ITTVPRSEEVQRN-----EVMAITPQSATPRGAIHLKRNCGDOPDCGVALAKTGEV 930
Db 1210 IGHQRPSPVIRPNRPANDTLAFAFPSCOI--SAFHQLAEIENHR-VVAVALPRCPBL 1266
Qy 931 FCGRAKINVELAEPATPKPRHAFQ-EGVQWYKVTNASKHQALOTLSRYTKRSADLP 989
Db 1267 EOGLLYLPOELTTCDSVV---TFELTDIVHCRMAFQSKKAVLSTLVGRYGGRT---XL 1319
Qy 990 HEAK-EDYKRMNLNSIDRHMDVTVEDABRAVFEQTKTORGTEDELLPDD--DPYIR 1046
Db 1320 YNASHSDVR---DSLARFIPAIGPVQVTTCEIYELVEANVEKGGDSAVLEIDLCTRDV 1376
Qy 1047 DIDPLAKTOOK-VSPKINTGVGOGIAAHSKSLNVLAAWIRILEE-ILRTGSRVRY 1104
Db 1377 RITFPQKCNKFTTGTIAGKVGOGISAMSKTFCALPGWPFAIEKAILALLPGVFG 1436
Qy 1105 NGLPDEE-EAMLEAKINOVPIATFVSADWTEPDTAANNSTBLLFPAALLERIGTPAAVN 1163
Db 1437 DAFDDTVFSAVAALAAKASVY---FENDFSESDSTONNFSGLBECAIMBECMPMLIR 1491
Qy 1164 LFERRCGRKTLRAKIGSVVNDGL-----LDGAAATPCRNITFSAWMLT--FRGVKF 1216
Db 1492 LY-----HLIRSAWMLQAPKESLRFMKHSGSEPGTLLMNTVWNAVAVTHCYDFDFDPOV 1545
Qy 1217 AAFKGDSDL-LGSHYLRFDASRLHMGERYKTGLKVEYOKIVPIGLVLSAEQVVL-DP 1274
Db 1546 AAFKGDSDIVLCSERYQSPGAANVLINGCGLK---LVVDRPPIGLVGVVVAPEGLALP 1602
Qy 1275 VRS 1278
Db 1603 VRS 1606

```

```

RESULT 12
US-09-553-427-7
; Sequence 7, Application US/09553427
; Patent No. 6379891
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R

```

```

APPLICANT: Yarbough, Patrice O
APPLICANT: Bradley, Daniel W
APPLICANT: Kravczynski, Krzysztof Z
APPLICANT: Tam, Albert
APPLICANT: Fry, Kirk B
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
NUMBER OF INVENTION: No. 6379891-A/No. 6379891-B Hepatitis Viral Agent
CORRESPONDENCE ADDRESS:
ADDRESS: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/553,427
FILING DATE:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/478,507
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US/08/279,823
FILING DATE: 25-JUL-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/681,078
FILING DATE: 05-APR-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/505,888
FILING DATE: 05-APR-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/420,921
FILING DATE: 13-OCT-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/367,486
FILING DATE: 16-JUN-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/336,672
FILING DATE: 11-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Sholtz, Charles K.
REGISTRATION NUMBER: 38,615
REFERENCE/DOCKET NUMBER: 4600-0183.22
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1693 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-09-553-427-7
Query Match 5.4%; Score 479; DB 4; Length 1693;
Best Local Similarity 21.2%; Pred. No. 1,4e-32;
Matches 369; Conservative 188; Mismatches 577; Indels 610; Gaps 73;

```

QY 118 RDGSRHITALE-----SRSVATGRBEPKADSLANGLASRTFCVDSGCAFKSRVG 170
 Db 113 RDVGRWYTAFTPRGRANCRSALRGLP-----AADRTGLDGSFGCNFRPRTG 160
 QY 171 IANHSIYDVTLEELANAFENHGLHMTAFAMMBELLIMDNVNAELGYRFAVIEBPMV 230
 Db 161 IALYSIHDMSPSDVAEMFRHGMTRLYAALHLPPEVLL-----PGTY 203
 QY 231 KDCAPGGDLRLHPELDFINESOERRIERLAAGSISRRAVIFSGDDMDGAYLHD-- 287
 Db 204 RTASVY-----LLIH-----DGRV-----VVTYEGDTAG--YNHDSN 235
 QY 288 FHTMLAVLVNRYPTPFGSLHIEVORRHSSIELRTTRAP-----FGDMLAVPR 339
 Db 236 LRSM-----IRTKVGDHPLIVERVAIGCHPULLITAPEPSMPYVYPRSTEVYR 290
 QY 340 TSGGLCRIPNIFYADASGTEHKITLTSQHKVNMLNFMQTRPKELVDMTVLMSFAR 399
 Db 291 SIFPGGTPELF--PTSCSTKSTFHAVPAHIMRLMFGATLDDQAF--CSRLMTYLRGI 347
 QY 400 LRAIVASEVTESWNISPADLVRTVSLV-----LHIEERRAAVAV 443
 Db 348 SYKVITGLVANEGMNASDALTAVITAAVLTICHORYLRTQAIKGMRLREHHAQFI 407
 QY 444 -----KTAQDVFG--ETSFMESLKHVLS----- 466
 Db 408 TRISWLFKSGSDYIPGRQLEFYAOCRRWLSAGFHLDRVLVFPDESAPCHRTAIRKAL 467
 QY 467 ---CC-----GLRNKGTDVYFTKR 483
 Db 468 SKFCCEMKWLGQECTCFLQPAEGAVGDQHDNEAEGSDVDPASASIDISGVVVGTA 527
 QY 484 VVDKRVHSL-----GRICDVR-----SP 504
 Db 528 LQPLDALDLPAEIVARAGRLTATVKVSQVGRIDCETLLGNKTFRTSPVDGAVLETNGP 587
 QY 505 EQ-----VGFLPSRVP----- 515
 Db 588 ERINLSFDASQSTMAAGPSFLTYAASAGLEVRYVAAGLDHRAVFAFGVPRSAFGVTA 647
 QY 516 -----PARVPHD----- 522
 Db 648 FCSALYFNRBAQRHSLIGNLMFHPBGLIGLPAFPSGHWESANPFCGSTLYTRTWSE 707
 QY 523 -----REBELVLEAG-----CYNERPVPSTPVEPOG----- 551
 Db 708 VDAVSSPARDLGFMSEBPSIPSRATPTLLAPPBPDPSPSAPALAEPAASGATAGA 767
 QY 552 -----FDADLMHAT----- 560
 Db 768 PAITHQTAHRRLFTYPPGSKVPFAGSLFESTCTMLVNASNVDRHPOGGLCHAFYQRYPA 827
 QY 561 ---AASL-----PRYR-----ATLOAG-----LN 576
 Db 828 SFDAAAFVWRDGAATAVLTGPRPIIHAAVAPDYRLHNPKRLBAAYRETCSRLGTAAYVLLG 887
 QY 577 TDVKQ-----LKIT 585
 Db 888 TGIYQVPIGPSFDMERNHRPGDELVLPELAAMFEANRPTRPTLTITTEDVARTANLAI 947
 QY 586 LBNAL---KTIDGLTSPVGLMEVY---GPPSGKTGTLIALEAAGKALVYAPREL 639
 Db 948 LBSATDVGRACACRVTY--GVOYQOTTAGVPSGSRSTOA-----DVDVVVVPREL 1000
 QY 640 REAMDRRIKPPSASATQVALAIL--RRATAEGAPFATVVIDECFMPFLVYVAIVHALSP 697
 Db 1001 RNAMRRR---GPAFPTHATAVYTGRR-----VIDEAPSLP--PHLLH--MQR 1045
 QY 698 SSRIVLVGDVHQGFIDFQSTSAMPLVRYVVKOCRRRTNQTCKRCADVAVATTFPSQSLY 757
 Db 1046 AATVHLGLDGNQIPALIDFEHAGL--VPAIRPDLGP--TSMNHVTHRWADV--CELIRGAY 1100
 QY 758 PGCTTTSQVAAISHVAPDYRNSQAOTLCTJOEKSNGHAGAMTYHEAGRTFASVILH 817

Db 1101 PMIQTTSRVLRSLEWGP-----AVGQKLVTOQAXPAN--PGSVTYHEAQAATYTERIT 1154
 QY 818 YNSTABQKLLAERK--HLIVGITRTNHLIYEDPPGIDERQOLNHSKAEVFTDIPAPLE 875
 Db 1155 ---ATADARGLQSSBAHAIVALTHTRETKCVIIDAIG--LIRBGIS--DALVNNPFLAGGE 1209
 QY 876 ITTVKPSSEVQRY-----EVMATIPPOSATPHGAIHLIRKNPDPDPOCGVALAKGYEV 930
 Db 1210 IGHQBSVLPKGNPDANDTILAPPPSCQI--SAFHQLEBELGHRP--VPAALVPPCPEL 1266
 QY 931 FGGRAKINVELAPDAPTPKHPAQ--EGVQWVYVYVNASNGHQAOTLLSRYYTRGADLPL 989
 Db 1267 EQLLYLPQGLTTCDSV-----TFBELTDIYHCSMAAPSGOKAVLSTLVGRYGRT---KL 1319
 QY 990 HEAK--EDVYKMLNSLDRHMDMTYEDARDRAVETQKTPQRTGYEDLLEPD--DPYIR 1046
 Db 1320 YNASHSDVR---DSLRFIPIALGPVQVTTCELYELVEAMVEKQDGSAYLELDLCNRDVS 1376
 QY 1047 DIDFLMKTOOK--VSPKINTGKVGQIIAHSKSLNPLVLAWIRILBE--LIRTSRYRYS 1104
 Db 1377 RITFPQDCKFTTGHTIAHGVGQISAMSKTFCALFGFWFAIERAIIALLPQGVFG 1436
 QY 1105 NGLPDEE--EAMLEAKINQVPATFVSADWTBPDTAHNTSEBLLPALLERIGTPAAVN 1163
 Db 1437 DAFDVTVPSAAVAARASNV-----FENDPSEBDSYONNPSLGBECALMEBCMPWLIR 1491
 QY 1164 LFRERCGKTLRAKGLSVVVDGL-----LDGAMTTPCRRNTIFSAAVMTL--FRGVKF 1216
 Db 1492 LY-----HLIRSAWILQAPKESLRGFWKHSGBEPGTLNMTVMNNAVITHCYDFRDFQV 1545
 QY 1217 AAFKGDSTL--LQSHYLRPDASRLHNGERKTHGLVYEVQKTYPIYGLVSAQVUL--DP 1274
 Db 1546 AAFKGDSTVLCSEYHQSREAAVLIAGCGLK--LVNDFPIGLYAGVVAPGLGALPVD 1602
 QY 1275 VRSA 1278
 Db 1603 VRPA 1606

RESULT 13
 US-08-093-453B-2
 ; Sequence 2, Application US/08093453B
 ; Patent No. 5439814
 ; GENERAL INFORMATION:
 ; APPLICANT: Frey, Teryl K.
 ; APPLICANT: Dominguez, Geraldina
 ; APPLICANT: Wang, Chin Yen
 ; TITLE OF INVENTION: Modified Infectious Rubella Virus
 ; NUMBER OF SEQUENCES: 13
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSER: Jamie L. Greene, Jones & Askew
 ; STREET: 191 Peachtree Street, 37th floor
 ; CITY: Atlanta
 ; STATE: Georgia
 ; COUNTRY: United States
 ; ZIP: 30303
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50
 ; COMPUTER: Macintosh
 ; OPERATING SYSTEM: 7.0
 ; SOFTWARE: Microsoft Word
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/093,453B
 ; FILING DATE: 19 JUL 1993
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: U.S. 07/722,334
 ; FILING DATE: 28 JUN 1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Greene, Jamie L.
 ; REGISTRATION NUMBER: 32,467
 ; REFERENCE/DOCKET NUMBER: 07362-0101


```

Db 531 VEDGCEPEVEAKCESEVIVDPSSSEPPVOEVLSTNGVQAARTEBEVGGDTGAGVAKSEV 590
Qy 166 KSRVGINHSLDYDTLE-----ELANAFENHGLHWRAFMH 202
Db 591 SQRVFPQVPAHEAGLEASSGAVVEPLQVSVPAVEKTVLSVEKARLKAVDKAKAVHA 650
Qy 203 PELLVMDNVNNAELGRFHVIEBPMAVKD-CAFOGDLRLHPELDPINESQERRIERL 261
Db 651 KE-----VKQVVKTLPRGALKISBDTVRKELCMFRTSCGV-QLDVYNEA-----TI 697
Qy 262 AARGSVS-----RRAVIFSGDDDMGDAY-----LHDPHTMLAY--- 294
Db 698 ATRFSNAFTVDSLKGRSAVFPFS---KLGEQTYNGSHVSSGMPRALBEDILTAKTSPV 754
Qy 295 ---LVVNYPTPGFSLHIEVQRHSSIELRI-----TRAPGDBMLAVPRTS 341
Db 755 FDHCLVQKYMGGGVPPHADDEBCYPSDNPILTVNLVKANFSTCKRKGKVMYI----- 809
Qy 342 QGLCRINIFYYADASTEHKTLITSOHKVNL-----LNFQOTR----- 381
Db 810 ---NVASGDYFLMPCGFQR---THLHVSNSIDEGRISLTFRATRREVFVGRLQLAG 861
Qy 382 -----PEKELVDMTVL-----MSF 395
Db 862 VSDKSPGVBNQOPOGQATRTTPKSGGKALSSEGREVKRSTYSIWCEQDYVRKCEW 921
Qy 396 ABAARLAIIVASVEETSESNISPADIVRTVSLVYLIH---IEERRAA-----VAKTA 446
Db 922 LRADNPVMAKPGYTPMTEFVVKAGTSEDAVEELKTLAIGITRYALLMARNIAVTTA 981
Qy 447 KDDVFGETSFWESLK--HVLGSCCGRLNLTGTDVFTKRVVDKRVHSLGDIICDVRISP 504
Db 982 EGVIAKVNQYVESLPGFHVYKS-----CTDLIF-----HSTQDGL----- 1016
Qy 505 EOVGFLEPRVPAVFPDREBLEVLRAAGCYNERPVPSTPVEBPQFDALMHAATAASL 564
Db 1017 -RVADLP-----YVF-----IAEKGIIF-----IKGDVVD----- 1039
Qy 565 PEVRATIQAGINTDVOKKITLENALKITIDGLTISPVRGL-----EMYEGPPSG 614
Db 1040 ---AAVALGDNLSVCDLILFHDAILNLMGALKVARC-GWGESFKSFEYKCYNAPPGG 1094
Qy 615 KTGLTIALBAAGSKALYVAPTRRELRAMPDRIRKPSASAT-----QHVALALIR- 665
Db 1095 KTMVLV-----DEFVSPNSTATITANVGSSEBINNAVKRD 1131
Qy 666 ATAGABFAT-----VVIDECFMPFLVVAIVHALSPSSRIVLVDVH 708
Db 1132 PNLEGLMSATTVNSRVVNFIVRGMYKVLVDEVYMMHQLQLGVPATGASEGLFPGDIN 1191
Qy 709 QIGFID-----FOGTSANMPLYRVDVVKQCRRTFNOTKRC-PADV 747
Db 1192 QIPIINEKVFPRMDCAVFPKRESVYTSKSYRCPL--DVCYLLSSMTVRTEKCYBEKV 1249
Qy 748 VA---TTFOSLYPGCTTSGCVASISHVADYRNSOQTLCTFOEKS--RHGAEG-- 799
Db 1250 VSGDKRVVNSLSRPIGTIDDAEIN--ADVY-----LCMTQLEKSDMKRSLKGGK 1300
Qy 800 ---AMTVEAOGRTFASVILHYNSTABOKLLAKRSHLVGITHTNHL----- 845
Db 1301 ETPTMTVHAOGKTFSDVVL-FRTKKADDSLFTKQPHILVGLSHSTSLVVAALSSELD 1359
Qy 846 ---YIRD--PTGDIERQLNHSAAQAEVFTDI 870
Db 1360 KVGTYISDASPOSVDALHLTFAPAGCFRGI 1390

```

RESULT 15
 US-09-579-259-2
 ; Sequence 2, Application US/09579259
 ; Patent No. 6558953
 ; GENERAL INFORMATION:
 ; APPLICANT: Goncalves, Dennis

```

; Ling, Kai-Shu
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS
; PROTEINS AND THEIR USES
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/579,259
; FILING DATE: 25-May-2000
; CLASSIFICATION: <unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60009008
; FILING DATE: 21-DEC-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/621
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1600
; TELEFAX: (716) 263-1304
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1390 amino acids
; TYPE: amino acid
; STRANDEDNESS: <Unknown>
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
; US-09-579-259-2

Query Match 2.44; Score 213; DB 4; Length 1390;
Best Local Similarity 20.34; Pred. No. 4.9e-09;
Matches 213; Conservative 118; Mismatches 326; Indels 394; Gaps 50;

Qy 60 QHPLAGHGVABEVLNFPARGSTYLR---TQPSLSALKLHGAAPVADYHGTCTYKG 116
Db 494 EHP-ACGRABEVEDLITPLG---TAVLESPPVGBRAGS-----APN----- 530
Qy 117 TRDSRHITLESRSVATGRPEPRKADASLL---ANGI-ASRT-----FCVDGVSGCAF 165
Db 531 VEDGCEPEVEAKCESEVIVDPSSSEPPVOEVLSTNGVQAARTEBEVGGDTGAGVAKSEV 590
Qy 166 KSRVGINHSLDYDTLE-----ELANAFENHGLHWRAFMH 202
Db 591 SQRVFPQVPAHEAGLEASSGAVVEPLQVSVPAVEKTVLSVEKARLKAVDKAKAVHA 650
Qy 203 PELLVMDNVNNAELGRFHVIEBPMAVKD-CAFOGDLRLHPELDPINESQERRIERL 261
Db 651 KE-----VKQVVKTLPRGALKISBDTVRKELCMFRTSCGV-QLDVYNEA-----TI 697
Qy 262 AARGSVS-----RRAVIFSGDDDMGDAY-----LHDPHTMLAY--- 294
Db 698 ATRFSNAFTVDSLKGRSAVFPFS---KLGEQTYNGSHVSSGMPRALBEDILTAKTSPV 754
Qy 295 ---LVVNYPTPGFSLHIEVQRHSSIELRI-----TRAPGDBMLAVPRTS 341
Db 755 FDHCLVQKYMGGGVPPHADDEBCYPSDNPILTVNLVKANFSTCKRKGKVMYI----- 809
Qy 342 QGLCRINIFYYADASTEHKTLITSOHKVNL-----LNFQOTR----- 381
Db 810 ---NVASGDYFLMPCGFQR---THLHVSNSIDEGRISLTFRATRREVFVGRLQLAG 861
Qy 382 -----PEKELVDMTVL-----MSF 395

```

```
Db      862 VSDKSPGVNPGQPSQGRTRITTPKSGKALSGSGREVKRSTYSICEDDYVRKCEM 921
Qy      396 AARLRAIVASEVTESSWNISPADLVRTVSLVYLHI---IERRPA-----VAVKTA 446
Db      922 LRADNPVMAIKPGYPMTEFVYKAGTSBDVAVVEYLKYLAIIGRTYRALLMARINIAVTTA 981
Qy      447 KDDVFGETSFWESLK--HYLGSCCGIARNIKGTDVVFTRKVDKVRHSLGDIICDVRISP 504
Db      982 EGVLVKPNQVYSLPGFHYKKS-----GTDLIIF-----HSTODGL----- 1016
Qy      505 EGVGFPSRVPARVPHDREBLEVLRACYNRPPVSTPVEEPQGFADLMHATPAASL 564
Db      1017 -RVRLP-----YVF-----IAEKIF-----IKGDVD----- 1039
Qy      565 PEYRATLQAGLNTDVNQLKITLENALKTIDGLTSPVRL-----EMVEGPPSG 614
Db      1040 ----AVVALGDNLVSCDDILVFHDAINLMLKVARC-GMVGESFKSFEYKCYNAPGCG 1094
Qy      615 KTGTLIALEAAGKALVYAPTRLEAMDRRIKPPASAT-----QHVALAILRR- 665
Db      1095 KTTMLV-----DEFKSPNSTATITANYGSSSEDINMAVKKRD 1131
Qy      666 ATAEGAPFAT-----VYIDECFMFPLVYVAIVHALSPSRIVLVGDVH 708
Db      1132 PNLBGINSATTVNSRVNFIVRGMVYKVALDEVYMMHOGILQGVATGASBGLFFGDIN 1191
Qy      709 QIGFID-----FOGTSANMPLVRDVVKQCRRTFNQTKRC-PADV 747
Db      1192 QIPFIRREKVFNRMDCAVFPVKESVYVTSKSYRCPL--DVCYLLSMTVRGTEKCYPEKV 1249
Qy      748 VA---TTFQSLYPCCTTSGCVASISHVAPDYRNSQAOTLCFTQEEKS--RHGAG--- 799
Db      1250 VSGKDKPVVRSLSKRPDGTDDVAEIN--ADVY-----LCMTQLEKSDMKRSLKXGK 1300
Qy      800 ---AMTVHEAQGRTPASVILHYNGSTAEQKLAEKSHLVGIFRHTNHL----- 845
Db      1301 ETPVMTVHEAQGRTPASVILHYNGSTAEQKLAEKSHLVGIFRHTNHL----- 845
Qy      846 ---YIRD--PTGDIERQLNHSAKAEVFTDI 870
Db      1360 KVGTYISDASPGVSDALHTFAPAGCFRGI 1390
```

Search completed: January 15, 2004, 16:54:03
Job time : 50.9919 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 16:52:13 ; Search time 204.393 Seconds

(Without alignments)
1704.735 Million cell updates/sec

Title: US-09-991-262-40

Perfect score: 8832

Sequence: 1 MVAKATDVAAVAAADVAA.....LKRIRSDVDPVKISKSA 1704

Scoring table: BLOSUM62

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

1: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*

2: /cgn2_6/ptodata/1/pubppa/US06_PUB_PUB.pep:*

3: /cgn2_6/ptodata/1/pubppa/US06_PUB_PUB.pep:*

4: /cgn2_6/ptodata/1/pubppa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubppa/US07_PUB_PUB.pep:*

6: /cgn2_6/ptodata/1/pubppa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubppa/US08_PUB_PUB.pep:*

8: /cgn2_6/ptodata/1/pubppa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*

10: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubppa/US09_PUBCOMB.pep:*

12: /cgn2_6/ptodata/1/pubppa/US09_PUB_PUB.pep:*

13: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*

14: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*

15: /cgn2_6/ptodata/1/pubppa/US10_PUBCOMB.pep:*

16: /cgn2_6/ptodata/1/pubppa/US10_PUB_PUB.pep:*

17: /cgn2_6/ptodata/1/pubppa/US60_PUB_PUB.pep:*

18: /cgn2_6/ptodata/1/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	8832	100.0	1704	11	US-09-991-262-40
2	515	5.8	1698	11	US-09-468-147-91
3	512.5	5.8	1708	11	US-09-468-147-166
4	479	5.4	1693	11	US-09-851-410-7
5	294	3.3	1694	15	US-10-223-070-19
6	253.5	2.9	2116	12	US-10-224-999A-3475
7	228.5	2.6	961	15	US-10-223-070-4
8	213.5	2.4	1707	12	US-10-388-848-2
9	213	2.4	1390	12	US-10-138-842A-2
10	213	2.4	1390	12	US-10-039-112-2
11	205	2.3	1236	11	US-09-769-787-109
12	198	2.2	431	11	US-09-851-410-2
13	191.5	2.2	1187	15	US-10-223-070-23
14	186.5	2.1	1367	10	US-09-801-368-108
15	186.5	2.1	2478	9	US-09-815-242-5816

16	186.5	2.1	2478	9	US-09-815-242-12967
17	183	2.1	844	15	US-10-156-761-7653
18	182.5	2.1	3262	12	US-10-379-381-4
19	178.5	2.0	2109	15	US-10-223-070-6
20	176	2.0	19723	12	US-10-084-846A-5
21	173.5	2.0	19725	12	US-10-084-846A-4
22	172	1.9	5877	15	US-10-142-515-11
23	170.5	1.9	725	12	US-10-369-493-2004
24	170	1.9	1139	15	US-10-223-070-11
25	169	1.9	300	12	US-10-138-842A-50
26	169	1.9	2429	12	US-10-288-798-8
27	169	1.9	19695	12	US-10-084-846A-3
28	168.5	1.9	797	15	US-10-156-761-10907
29	167.5	1.9	1462	10	US-09-971-536-69
30	167	1.9	5935	15	US-10-243-243A-8
31	166.5	1.9	980	12	US-10-029-386-33686
32	163.5	1.9	1060	10	US-09-955-909-2
33	163.5	1.9	2368	9	US-09-815-242-5635
34	163.5	1.9	2368	9	US-09-815-242-12389
35	163	1.8	630	15	US-10-207-655-69
36	163	1.8	1852	12	US-10-023-649-2
37	163	1.8	3268	12	US-10-379-381-2
38	162.5	1.8	1794	12	US-09-965-738-299
39	162.5	1.8	1799	12	US-09-965-738-149
40	162.5	1.8	1821	12	US-09-965-738-82
41	162.5	1.8	11721	12	US-09-965-738-162
42	161.5	1.8	1162	10	US-09-745-008-34
43	160.5	1.8	483	15	US-10-029-840-4
44	160.5	1.8	2231	12	US-10-379-381-5
45	157	1.8	258	12	US-10-104-047-3034

ALIGNMENTS

RESULT 1

US-09-991-262-40

Sequence 40, Application US/09991262

Publication No. US20030041349A1

GENERAL INFORMATION:

APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.

TITLE OF INVENTION: Insect Viruses and Their Uses in Protecting Plants

NUMBER OF SEQUENCES: 57

CORRESPONDENCE ADDRESSES:

ADDRESS: Flehr Hobach Test Albritton & Herbert LLP

STREET: Four Embarcadero Center, Suite 3400

CITY: San Francisco

STATE: California

COUNTRY: United States

ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/991,262

FILING DATE: 20-NOV-2003

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/234,238

FILING DATE: 20-JAN-1999

APPLICATION NUMBER: US 08/485,355

FILING DATE: 07-JUN-1995

APPLICATION NUMBER: US 08/440,522

FILING DATE: 12-MAY-1995

APPLICATION NUMBER: US 08/089,372

FILING DATE: 08-JUL-1993

APPLICATION NUMBER: AU PL4081/92

FILING DATE: 14-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Trecartin, Richard F.

Sequence 12967, A

Sequence 7663, Ap

Sequence 4, Appl1

Sequence 6, Appl1

Sequence 5, Appl1

Sequence 4, Appl1

Sequence 11, Appl1

Sequence 2004, Ap

Sequence 11, Appl1

Sequence 50, Appl1

Sequence 8, Appl1

Sequence 3, Appl1

Sequence 10907, A

Sequence 69, Appl1

Sequence 8, Appl1

Sequence 33686, A

Sequence 2, Appl1

Sequence 5635, Ap

Sequence 1389, A

Sequence 69, Appl1

Sequence 2, Appl1

Sequence 299, Appl1

Sequence 82, Appl1

Sequence 162, Appl1

Sequence 34, Appl1

Sequence 4, Appl1

Sequence 5, Appl1

Sequence 3034, Ap

REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-58631-3/RFT/DSS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELEX: 910 277299
 INFORMATION FOR SEQ ID NO: 40:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1704 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 40:
 US-09-991-262-40

Query Match 100.0%; Score 8632; DB 11; Length 1704;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1704; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MYAKTIDVARVVAADVAAYANLQORAVKLDPAAPLKALETLRLYYPLRFKGTLPPTQ 60
 DB 1 MYAKTIDVARVVAADVAAYANLQORAVKLDPAAPLKALETLRLYYPLRFKGTLPPTQ 60
 QY 61 HPILAGHORVAEEVHLNFAKGRSTYLEIGPSLSALKLHGAPNAPADYHGGCTKGTDRG 120
 DB 61 HPILAGHORVAEEVHLNFAKGRSTYLEIGPSLSALKLHGAPNAPADYHGGCTKGTDRG 120
 QY 121 SRHITALESRVAVTRPERKADASLLANGIASRTFCVDVGSCAPKRSRGVIANHSLYDT 180
 DB 121 SRHITALESRVAVTRPERKADASLLANGIASRTFCVDVGSCAPKRSRGVIANHSLYDT 180
 QY 181 LEELANAEENHGLHNVRAFMHMBELLVYDNVYNAELGYRFHYIEBPAAVXKCAFGGDL 240
 DB 181 LEELANAEENHGLHNVRAFMHMBELLVYDNVYNAELGYRFHYIEBPAAVXKCAFGGDL 240
 QY 241 RLHPFELDFINESOERIRLARGSYSRRAVIFSGDDMDGAYLHDFHTMLAYLLVRY 300
 DB 241 RLHPFELDFINESOERIRLARGSYSRRAVIFSGDDMDGAYLHDFHTMLAYLLVRY 300
 QY 301 PTFPGSLHIEVQRHSGSIELRTAPPGDBMLAVPRTSQGLCPNIFYADASGTE 360
 DB 301 PTFPGSLHIEVQRHSGSIELRTAPPGDBMLAVPRTSQGLCPNIFYADASGTE 360
 QY 361 HTTILTSQHKVMNLNFMQTRPEKELVDMTVLMSFARALRAIVASETTESSWINSPAD 420
 DB 361 HTTILTSQHKVMNLNFMQTRPEKELVDMTVLMSFARALRAIVASETTESSWINSPAD 420
 QY 421 LVRTVVSLLVHLIIEERRAAVAVKTAQDVPGETSPWESLKHVLSGCGRLNKGTDVYF 480
 DB 421 LVRTVVSLLVHLIIEERRAAVAVKTAQDVPGETSPWESLKHVLSGCGRLNKGTDVYF 480
 QY 481 TKRVVVKYRVHSLGDIICDVRLSPEQVGLPSKRVPAVFDHREBELVLRKAGCYNERPV 540
 DB 481 TKRVVVKYRVHSLGDIICDVRLSPEQVGLPSKRVPAVFDHREBELVLRKAGCYNERPV 540
 QY 541 PSTPVEBEQGFADALMHTAASLPYRATLQAGLNTDVYKQLKITLNNLKTITDGLTSP 600
 DB 541 PSTPVEBEQGFADALMHTAASLPYRATLQAGLNTDVYKQLKITLNNLKTITDGLTSP 600
 QY 601 VAGLEMEYEGPSSGKTGTLIALLEAAGKALVYAPTRERLREAMDRIKPPSASATQVAL 660
 DB 601 VAGLEMEYEGPSSGKTGTLIALLEAAGKALVYAPTRERLREAMDRIKPPSASATQVAL 660
 QY 661 ALIRBATAEGAPFATVIDECFMPFLVYVAIVHALSPSSRIYLVGDVHIGIFDFQGTSA 720
 DB 661 ALIRBATAEGAPFATVIDECFMPFLVYVAIVHALSPSSRIYLVGDVHIGIFDFQGTSA 720
 QY 721 NMPPLVRDVVKQCRRRRFNQTGCPADVYATTPFQSLYPCGTTSSGCVASISVAPRYRS 780
 DB 721 NMPPLVRDVVKQCRRRRFNQTGCPADVYATTPFQSLYPCGTTSSGCVASISVAPRYRS 780
 QY 781 QAOTLCFTQEKSRHGAEGAMTVHEAGRTFASVLIHNGSTAEQCLAEKSHLVGTR 840

DB 781 QAOTLCFTQEKSRHGAEGAMTVHEAGRTFASVLIHNGSTAEQCLAEKSHLVGTR 840
 QY 841 HTNHLIYRDPGDIERQINSHAKAEVFTDIPAPLEITTVKPSSEYORNEVMAITPPQAT 900
 DB 841 HTNHLIYRDPGDIERQINSHAKAEVFTDIPAPLEITTVKPSSEYORNEVMAITPPQAT 900
 QY 901 PHGATHILRKQPGQPDGCVALKATGVEVGGRAKIVNVEASBPATKPRHAPQEGVQW 960
 DB 901 PHGATHILRKQPGQPDGCVALKATGVEVGGRAKIVNVEASBPATKPRHAPQEGVQW 960
 QY 961 VKVTNASHKHOALQTLISRYTKRSADLPLHAKEDVKMLNSLDHMDMTYTEDARDRAV 1020
 DB 961 VKVTNASHKHOALQTLISRYTKRSADLPLHAKEDVKMLNSLDHMDMTYTEDARDRAV 1020
 QY 1021 FETQKFTQRGQGVEDLLEPDDPYRIDIDPLMKTQOKVSPKINTGVGGIJAHSKSLN 1080
 DB 1021 FETQKFTQRGQGVEDLLEPDDPYRIDIDPLMKTQOKVSPKINTGVGGIJAHSKSLN 1080
 QY 1081 FYLAAMIRILBELILRTGSRTRYNSGLPDEBEAMLEAKINOVPHATVPSADMTSEFDTH 1140
 DB 1081 FYLAAMIRILBELILRTGSRTRYNSGLPDEBEAMLEAKINOVPHATVPSADMTSEFDTH 1140
 QY 1141 NNTSELPAALLERIGTPAALAVNLFRRCGKRTLRKAGLSYEVVDGLDSCGAAMPCCNT 1200
 DB 1141 NNTSELPAALLERIGTPAALAVNLFRRCGKRTLRKAGLSYEVVDGLDSCGAAMPCCNT 1200
 QY 1201 IFSAAVMLTLFRGVKPAAPKDDSLCGSHYLRPDASRLHNGERYKTHLYEVOKIVPY 1260
 DB 1201 IFSAAVMLTLFRGVKPAAPKDDSLCGSHYLRPDASRLHNGERYKTHLYEVOKIVPY 1260
 QY 1261 IGLVSAKOVYLDPRSALKI FGRCTSELLYSKYVEAVRDTTKMSDARVHSLCHMSA 1320
 DB 1261 IGLVSAKOVYLDPRSALKI FGRCTSELLYSKYVEAVRDTTKMSDARVHSLCHMSA 1320
 QY 1321 CYNTAPESAAVYIIDAVERFGSDPFFEQLRVRAHVOAPDAYSSTYPANVYASCLDHYF 1380
 DB 1321 CYNTAPESAAVYIIDAVERFGSDPFFEQLRVRAHVOAPDAYSSTYPANVYASCLDHYF 1380
 QY 1381 EPRQAAAPAGFATCAKETPESLSLAKAGVSAITSHVATGTAPPSPPMDAPANSFSELL 1440
 DB 1381 EPRQAAAPAGFATCAKETPESLSLAKAGVSAITSHVATGTAPPSPPMDAPANSFSELL 1440
 QY 1441 TPETPSTSSPSSSSSDSGRSLSGGDPARTTEDNSRKPQODQSRSSBCLDRSG 1500
 DB 1441 TPETPSTSSPSSSSSDSGRSLSGGDPARTTEDNSRKPQODQSRSSBCLDRSG 1500
 QY 1501 BRTGSSLTAPFAPSFPSEERARLATGPTVAAATSPSATPSCATDOVAARTTPDAPPL 1560
 DB 1501 BRTGSSLTAPFAPSFPSEERARLATGPTVAAATSPSATPSCATDOVAARTTPDAPPL 1560
 QY 1561 GSQSAVAVSKPYRPTTARWKEVTPPLHAWKGYTGDRPEYREDPTPAVVOALISGRYPQK 1620
 DB 1561 GSQSAVAVSKPYRPTTARWKEVTPPLHAWKGYTGDRPEYREDPTPAVVOALISGRYPQK 1620
 QY 1621 TKLSSDAKGYRTGCGOSTSPAPASADYQARDQTVRVCAAEAMASCHEPLASSA 1680
 DB 1621 TKLSSDAKGYRTGCGOSTSPAPASADYQARDQTVRVCAAEAMASCHEPLASSA 1680
 QY 1681 ASADLKRIKIRSTDSVPDYKISKSA 1704
 DB 1681 ASADLKRIKIRSTDSVPDYKISKSA 1704

RESULT 2
 US-09-466-147-91
 ; Sequence 91, Application US/09468147A
 ; Publication No. US20030049601A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Abbott Laboratories
 ; APPLICANT: Schleuder, George G.
 ; APPLICANT: Erker, James C.
 ; APPLICANT: Desai, Suresh M.
 ; APPLICANT: Dawson, George J.

APPLICANT: Mushahwar, I. K.
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING
 TITLE OF INVENTION: HEPATITIS E VIRUS
 FILE REFERENCE: 6232 US.P1
 CURRENT APPLICATION NUMBER: US/09/468,147A
 CURRENT FILING DATE: 1999-12-21
 EARLIER APPLICATION NUMBER: US 09/173,141
 EARLIER FILING DATE: 1998-10-15
 EARLIER APPLICATION NUMBER: US 60/061,199
 EARLIER FILING DATE: 1997-10-15
 NUMBER OF SEQ ID NOS: 258
 SOFTWARE: FastSeq for Windows Version 3.0
 SEQ ID NO 91
 LENGTH: 1698
 TYPE: PRF
 ORGANISM: Hepatitis E Virus
 FEATURE:
 OTHER INFORMATION: Xaa = Unknown or Other at position 174
 OTHER INFORMATION: Xaa = Unknown or Other at position 363
 OTHER INFORMATION: Xaa = Unknown or Other at position 1088
 OTHER INFORMATION: Xaa = Unknown or Other at position 1131
 OTHER INFORMATION: Xaa = Unknown or Other at position 1217
 OTHER INFORMATION: Xaa = Unknown or Other at position 1389
 US-09-468-147-91

Query Match 5.84; Score 515; DB 11; Length 1698;

Best Local Similarity 21.78; Pred. No. 4.3e-32;

Matches 397; Conservative 179; Mismatches 553; Indels 700; Gaps 84;

QY 13 AAADVAANVLOQAVKLDFAFPAKLETT-LHRUYRP--LRFKQGTLPFQHPILAGHQ 68
 DB 13 AAASALANAVVR-----PFLSRVOTELINLMQROLVFPEVL--MNHPI---Q 59
 QY 69 RVAEEVLHNFARGES-TVLEIGPSLSALKLHGAPNAPVADYHGC-TKYGTGDSRHITA 126
 DB 60 RVINHEIQYCRADAGRCLEVGAPHS--INDNPV---LHRCFLRPVGRDVGWYSA 112
 QY 127 LE-----SRVATGRPEFKAADSLANGIASRTFCYGVGSCAFKSRVGIANHSLYD 179
 DB 113 PTRGPANCRSRALRGLRP-----ADRTYCPDGFSCRCAFAETGVALYSLHDL 160
 QY 180 TLEELANFENHGLMVARAFMHPEBELLYMDNVNNAELGYRFHYIEBPMAVKCAFQGGD 239
 DB 161 WPAVDAEMARHGTRLYAALHLPPEVLLPPTYHTTSLYLIH-----DGD 206
 QY 240 LRLHPPELDFINESQERRIERLAAGSYSRRAVIFSGDDWDGDAYLHDFHTMLAYL--- 295
 DB 207 -----RAVVTYEGDTSAG--YNHVSTILRAMIRTTK 235
 QY 296 LVNRYPTPPFGSLHIEVORRHGSSIELRITRAP-----PGDRMLAVPRTSQGLCRI 347
 DB 236 IVGQHP-----LVIERVRAIGCHFVLTLTAAPESPMPVYVPRSTEVYRSIFPGGS 289
 QY 348 PNIFYVADASGETEKTILTSQHKVNMMLNFMQTRPEKELVDMTVLMSFARALAIYVAS 407
 DB 290 PSLP--PSACSTKSTFAAVPVHIDRLMLFGATLDDQAF--CSRMTYLKIGSYKVYGA 346
 QY 408 EVTESWNISPADIVRTVLSLYLHIER-----BRAAVA----- 442
 DB 347 LVANEGMNASEDALTAIXITAYLVITICQRYLRTQALISGMKRLGVEHAKQITRLYSMLF 406
 QY 443 VKTAKADV-----FGSTFWES-----LKHVLS--CCGLR 471
 DB 407 EKSGRDYIPGRQLQFYAQCRRLWSAGFLHDPRLVLFDESVPVRCRTPFLKVAAGKFCFMR 466
 QY 472 NL-----KGTDV----- 478
 DB 467 WLGCCTCTLEPAGSLVGDHGHNEAYGSEVDPAEPHLDVSGTYAVHGHQLEALYAL 526
 QY 479 -----VFTKRVVDKRYV----- 490
 DB 527 NVPODIABARARLTATVETELVASPDLBCRTVLGNKTFRTTVVDGHLHLEANGPEQYVLSFD 586

QY 491 -----HSL-----GDI--ICD----- 499
 DB 587 ASRSMAGSHSLTYELTPAGLOVRISNGLDCTATPPGAGSAAFGSVAFTSALYRY 646
 QY 500 -----VLSPEQV-----GF----- 509
 DB 647 NRFTQBSLGLWMLHPEGLLGIPTPPSPGHIMESANPFOEGTLYTRTWSTGFSFSDS 706
 QY 510 -----LBSRVPAR---VFHDESLBYLREAGCYNERPVST----- 543
 DB 707 PPEAAMAPATPGLPHSTPPVSDIWLPPSSEFQV-----DAAPVPPAPDAGLPG 759
 QY 544 -----PVSEEP-----QGFADL--W----- 557
 DB 760 PVLTTPPPPPVHKSPSPSRNRLLYTPDCAKTVAGSLFESDCMLYNASNPGHRPG 819
 QY 558 -----HATNASLPEYRA-----TLQAG- 574
 DB 820 GGLCHAFYQRPFAFYPTBPIHREGLAAYTLTPPIIHAVA---PDRIVENPRLLEAY 876
 QY 575 -----LNTDVYQLKTL----- 586
 DB 877 RETCSRGTAAYPPLGSGIYQVPVSLSFDAWERNHREDELYTEPANNFPAKPAQPV 936
 QY 587 -----ENALKT-----ID-----GLTSPVRLMEYE---GPPSGKTGLIAL 623
 DB 937 LTTEEDARTANLALBLDAATEVGRACAGCTISB--GIYHQTAGVSGSKSHSI---- 990
 QY 624 EAAGKALYVAPTELEAMDRRIKPPSASATQVA--LAILRATAGAPFAVTVIDEC 681
 DB 991 -QGGDVVVVVPTRELANSWR--GFAFTPTHTARVITIGR-----VVIDEA 1036
 QY 682 FMPELVYVAIVHALSPSSRIYLVGDVHQIGFIDFGTSANMPLVRDVYKQCRRTFNQTK 741
 DB 1037 PSLP-PLHLLH--MQRASVYHLGDPQOIPALIDREHAGL--VPAIRP--ELAPTSMMXYTH 1091
 QY 742 RCPADVATTFPQSLYPCGCTTSGCVASISHVAPDYNSSQATLCFPOEBSRSGAGAM 801
 DB 1092 RCPADV--CELRIRAVYKIQTSRVLSLFWNEB---AIGQKLVYQAAANAN--PGAI 1143
 QY 802 TVHEAQRTFASVLIHNGSTABOKLAERKS--HLVGTITHTNLYIRDPYGIEROLN 859
 DB 1144 TVHEAQATFTETIT--ATADARGLIQSRRAIYALTRHTEKCYILDAFG--LHBEVG 1199
 QY 860 HS-----AKAEVTDIPAPLEITTVKSEEVQSN--EVMATIPQSATPHGAILLR 909
 DB 1200 ISDVIVANFFLAGGEVQHRP-----SVIPRGPDDNLGTLQAFPPSCQ--SAHQOLA 1251
 QY 910 KNFGDOPDCGVALAKGYEVFGRAKINVEIAPDAPTPKPHRAFO--BGVQWYKVTNASN 968
 DB 1252 BELGHRP-APVAAYLPPCPBELBQGLYMPQELTVSDV---LVFELTDIHCMAAPSQ 1306
 QY 969 KHALQTLTSLRYTGSADLPHE--AKEDYVRMLNSLRHMDWTYEDARDAVETOLKXF 1027
 DB 1307 KVALSTLVGRYGRKT--KLYPEAHSDVR---BSLRAPIETTPPVQATTCETELVELYAM 1360
 QY 1028 TORGETVEDLEPDP--DPYTRIDIDPLMKTOOK--VSPRINTGKVGOGIHAHSKSLNFYLA 1084
 DB 1361 VEKGQDQSAVLELDLCNRDYSRIFFQKCKMFTGTETIAGKVGOGISASKTFCALFG 1420
 QY 1085 AMITILE-ELIKTSRTRVYSNGLPDSEEMMLEAKINOVPHATFVGSADMTTEPTAHNT 1143
 DB 1421 PMFRAIKETILALPNIIFYG---DAYESVFAAASVAGSCKVFERDSEFSTONNF 1476
 QY 1144 SELFAALLERIGTPAAVNLFRERCGKRTLRA-----KGLSVYEVNDGLDGSAMTPCR 1198
 DB 1477 SLGLECYVMEBCGPPQULIRLYHVRSAWILQAKESLKGPKKH-----SGSPGTLW 1530
 QY 1199 NTIFSAAVMLTL--FRGVKFAAFKGDG--LLCGSHYLRFPASRLHMBGRYTKHLXYEVQ 1255
 DB 1531 NTVMNMAIIAHCEYFRDFRVAFAFGDDSVLTCSDYRSRBAALIIAGGALK--LKVDYR 1587
 QY 1256 KIVPYIGLIVASQVVL--DPIRSALKIFGCTYSBLLYSKTVAVENRDTITKMSDARYHSL 1314

Db	1588	PGLGAVVVVAPGLGLTPDVVFA	GR-LSE	1621
Qy	1315	LCHMSACTYVAPESAAYITIDAVRFGRC	1343	
Db	1622	-----GPERAEQLRLAVCDFLRC	1639	
RESULT 3				
	US-09-468-147-166			
	Sequence 166, Application US/09468147A			
	Publication No. US20030049601A1			
	GENERAL INFORMATION:			
	APPLICANT: Abbott Laboratories			
	APPLICANT: Schlader, George G.			
	APPLICANT: Desai, James C.			
	APPLICANT: Deseal, Suresh M.			
	APPLICANT: Dawson, George J.			
	APPLICANT: Mushahwar, I. K.			
	TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DETECTING			
	FILE OF INVENTION: HEPATITIS E VIRUS			
	FILE REFERENCE: 6232.US..P1			
	CURRENT APPLICATION NUMBER: US/09/468,147A			
	CURRENT FILING DATE: 1999-12-21			
	EARLIER APPLICATION NUMBER: US 09/173,141			
	EARLIER FILING DATE: 1998-10-15			
	EARLIER APPLICATION NUMBER: US 60/061,199			
	EARLIER FILING DATE: 1997-10-15			
	NUMBER OF SEQ ID NOS: 258			
	SOFTWARE: FastSeq for Windows Version 3.0			
	SEQ ID NO 166			
	LENGTH: 1708			
	TYPE: PR			
	ORGANISM: Hepatitis E Virus			
	FEATURE:			
	OTHER INFORMATION: Xaa = Unknown or Other at position 322			
	OTHER INFORMATION: Xaa = Unknown or Other at position 331			
	OTHER INFORMATION: Xaa = Unknown or Other at position 445			
	OTHER INFORMATION: Xaa = Unknown or Other at position 448			
	OTHER INFORMATION: Xaa = Unknown or Other at position 634			
	OTHER INFORMATION: Xaa = Unknown or Other at position 646			
	OTHER INFORMATION: Xaa = Unknown or Other at position 811			
	OTHER INFORMATION: Xaa = Unknown or Other at position 1553			
	OTHER INFORMATION: Xaa = Unknown or Other at position 1578			
	OTHER INFORMATION: Xaa = Unknown or Other at position 1691			
	US-09-468-147-166			
	Query Match			
	Best Local Similarity 21.64; Pred. No. 7e-32;			
	Matches 393; Conservative 186; Mismatches 557; Indels 687; Gaps 83;			
Qy	13	AAADVAANYLQGRVAVLDPAPPLKALET--LRLVYP--LRFKGGTLPTTQHPILAGHQ	68	
Db	22	AAASALANAVVR-----PFLSRVQTEILINLQRPQOLVRRPVL--WNHPI-----Q	68	
Qy	69	RVAEEVLNFRGRS-TVLEIGPSLSALKLGAAPNAPVADYHG-C-TKYGTRDGSRHITA	126	
Db	69	RVINHELOYCRRAAGRCLEVGAPRS--INDNPV---LHRCLRLRVGADVQRWYSA	121	
Qy	127	LE-----SRVATGPERFKADASLNGIASRTTCVDGSGCAFKSRVGIANSXYV	179	
Db	122	PTRGPAANCRRSALRGPP-----VDRITCFDGFSCAPAAETGVALYSLHDL	169	
Qy	180	TLEELANAFENHGLWVRAFPMHPELLYMDVNVNMLEGRFVIEEPPAVVXDCAFQGGD	239	
Db	170	WRADVAAMARHGHTRLXALHLRPREVLPPGT-----YHTTSY-----	208	
Qy	240	LRLHPELDFINESQERRIERLAARGSYRAVIFSGDDWDGDAYLHDPHTMLAYL----	295	
Db	209	LLIH-----DGNRAVVTYEGDTSAG--YHNDVSLIRAWIRTK	244	
Qy	236	LVRVYPPRPFGLSHIEVQRHGSIEILRTAP-----PQDRMLAVVPTSGQLCRI	347	
Db	245	IVDHP-----LVIERVALICHFVLLTLTAAPSPSPVYPRSTEVYRSIFGPGS	298	
Qy	348	PNIFTYADASGTEHKTILTISQKNNMLANFMOTREKSLVDMTULMSAPARLAIYVAS	407	
Db	299	PSLF--PSACSTKSTFHAVPVIWMXLMFGATLDDAFC--CSRLMTYLRGISYKVTGVA	355	
Qy	408	EYTESMNIISPADLVRTVSLVVLHIIERRAAVAV-----	443	
Db	356	LVANEGMNASBEDALTRVITAAVLTICHQRLFTQAIISGMRLVEYHAQKITLYSMLF	415	
Qy	444	KTAKDV-----FGETSFMES-----LKHYGS--CCGLR	471	
Db	416	EKSGRDYIFGRQLQFAQCRRLSAGFHLXPRXLVFDSDSVRCRCPTLAKYAGPCCMR	475	
Qy	472	NL-----KQTV-----	478	
Db	476	WLQECTCELEPAEGLVGDQDNBAVEGSEVDPAEPAHLDVSGTYAVHGLEALYAL	535	
Qy	479	-----VTKRVVDKRV-----	490	
Db	536	NVDHDIADASRLTATVLEIVASPRILCEKTVLGNTPRTTYVDGHLNANGPEEYVLSFD	595	
Qy	491	-----HSY-----GDI--ICD-----	499	
Db	596	ASRQSGASHLTTELTPAGLOVNISSNGDCTATPFXGGAASAPSEYVAFCSALYRY	655	
Qy	500	-----VRSPEGV-----GLPSRV-----	514	
Db	656	NRFQHSYLTGGLMLHPBGLGIPPPFSGHIWESANFCSEGLYTRTWGSGSPDS	715	
Qy	515	-----PPARVPH--DREBLEVLR--AGCNE--RPV	540	
Db	716	PPAAAPASAAAPGLPYPTPVSDIWLPPPEESHVDAASVSEPAGLSPVLTPP	775	
Qy	541	PTSPVYBSP-----	549	
Db	776	PPPPVPRKATSPPTTRRLTYTPDGAQVYAGSLXESDCMLVNASPHGRPGGCLHA	835	
Qy	550	-----QGFDA-----DLMHATASLPEYRA-----TLQAG-----	574	
Db	836	FYGRPEAFYSTEFIWBSGLAVTLTPRITIAVA--PDVRVEGNPRLEBAARBTCSR	892	
Qy	575	-----LNTDYKQIKTL-----EN	588	
Db	893	RGTAAVPLLGSIGYVPVSLSPDABERNRPGDELYTEPAAMPEANKRQAPALTIED	952	
Qy	589	ALKT-----ID-----GLTLPVRLGLENR-----GPPSGKGTTLALBAAGK	629	
Db	953	TARTALMLAELIDATEVGCAGCTISP--GIWHTQPRAGVPGSKSKSI-----QQSDV	1005	
Qy	630	ALVYAPTRBELREAMDRIRIPSPASATQVNA--LAILRATAGEAPFATVVIDECMPFLV	687	
Db	1006			

```

QY 975 TLRSYTKRSADLPLHE-AKEDVYKMLNSLDHMDWTVTEDARDRAVEFTQKFTQRCGT 1033
D 1323 TLVGRYKRT--KLIEAHSVDV--ESLARFPTTGPRAATTCCELYELVEANVEKGD 1376
QY 1034 VEDLEBD--DPYRIDIDPLMKTOOK-VSPKPIITGVGGGIAHSAKSLNFWLAAMRIL 1090
D 1377 GSAVLELDLGNROVSRTTFQKCKNKTGTETIAHGKVGGISMSKTFCLRGPMWRAI 1436
QY 1091 E-BILRTGSTRVYNSCLPDEBEAMLEAKINOVPHATFVSADWTERDTAHNNTSEILFA 1149
D 1437 EKELALLPPIFYG----DAYESVFAAASVAGSGCMVFENFSEFSDTONNFSLELC 1492
QY 1150 ALLEIRGTAAAVLPERGCKRTLR-----KGLSVEVDGLDGSAAWTPCRRITFSA 1204
D 1493 VVHEECGPMQMLRLHLVRSAMTLOAPKESLKGFWKCH-----SGEPGLTMMTWMM 1546
QY 1205 AVM--LTLFRGVKFAFKGDS--LLCGSHYLRFPDASHLNGERYKTKHLKVEVQKIVPYI 1261
D 1547 AIIAHCKEFDPRFAAFKGDSDVLCSDYRQXRMAALLGCGLK--LKYDIRPGLYA 1603
QY 1262 GLIVSAEQVVL-DPVSALKI FGCYTSSELLYSKYEA VBDITKWSADARYHSLCHMSA 1320
D 1604 GVVVAPGLGTLDPVVRFA--GR--LSE-----KNWGP----- 1631
QY 1321 CYNNVAPESAAYIIDAVRFRG 1343
D 1632 -----GERAEQLRLAVCDPLRG 1649

RESULT 4
US-09-851-410-7
Sequence 7, Application US/09851410
Publication No. US20030124510A1
GENERAL INFORMATION:
APPLICANT: Reyes, Gregory R
Yarbough, Patrice O
Bradley, Daniel W
Krawczynski, Krzysztof Z
Tam, Albert
Fry, Kirk B
TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
No. US20030124510A1-A/No. US20030124510A1-B Hepatitis Viral
NUMBER OF SEQUENCES: 20
CORRESPONDENCE ADDRESS:
ADDRESSER: Dehlinger & Associates
STREET: 350 Cambridge Avenue, Suite 250
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,410
FILING DATE: 07-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/128,275
FILING DATE: <Unknown>
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
APPLICATION NUMBER: US 07/208,997

```

```

; FILING DATE: 17-JUN-1988
; ATTORNEY/AGENT INFORMATION:
; NAME: Pelliccioli, Joanne R.
; REGISTRATION NUMBER: 42,995
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (650) 324-0880
; TELEFAX: (650) 324-0960
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1693 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-851-410-7

Query Match 5.4%; Score 479; DB 11; Length 1693;
Best Local Similarity 21.2%; Pred. No. 3.9e-29;
Matches 369; Conservative 188; Mismatches 577; Indels 610; Gaps 73;

QY 2 YAKATDVARYAADAAYANVILQORAVKLDPAFPALBTLRLYYR--LRFGKGTLPPT 59
D 6 FIAKPGITTAIBQALAAANSALANAVVRPFLSHQIIBILINLMQPROLVFRPEVW-W 63
QY 60 QHPITLAGHQVABEVLNFRKGRS-TVLEIGSELHSLKLGAPNAPVADYHGC-TRYGT 117
D 64 NHPI-----QVINHELELYCRASGRCLGAPRS---INDPNV-----VHRCLFAPVG 112
QY 118 RDGSRHITALE-----SRVATGRPEFKAASLLANGIASRTFCVDVGSCAPKSRVG 170
D 113 RDVQRMVITAPTRPANCRRSALRGLP-----AADRTYCLIDGSGCNFPALGT 160
QY 171 IANSLDYDTLEELANA FENHGLHVA RPFMEBELLYNDNVNAELGYRFAVIEBMAV 230
D 161 IALYSLHDSPSDVAEMFRGHGTRLYAALHLEPEVTL-----PGTY 203
QY 231 KDCAFQGGDLRLHFPLEIDPTNBSQERIERLARAGSYSRRAVYFSGDDMDGAYLHD-- 287
D 204 RTASV-----LLIH-----DGRVV-----VVTYEGDTSAG--YMHDSN 235
QY 288 FHTWLVLYVRNPTFPGRSLHTEVQRRHGSIELEIRTRAP-----RGDMILAVVPR 339
D 236 LRSW-----IRTKVTGDHPLVIERVRAIGCHEVLLITAPBSPMPYVYPSTEVYR 290
QY 340 TSOGLCRIPNIFYYADASGTEHNTIITSQHKVMLNFMQTRBEKELVDWTVLMSFAR 359
D 291 SIFGPGDTSLF--PTSCGTSKTFHAPVPAHIMWRMLFGATLDDQAF-C-SRLMTYIRGI 347
QY 400 LRAIVVASBYTSSWNISPADLVRTVSLV-----LHIERRAAV 443
D 348 SYKVTGTIVANEGMNASBDALPAVITAAVLTICHORYLTAQISKMRRLBREHAKFI 407
QY 444 -----KTAQDVFG--BITSPWESLKHVIGS----- 466
D 408 TRLYSWLFKSGRDYIPGQLEFYAOCRRWLSGPHLDPRVLVFDSSAPCHCTAIRKAL 467
QY 467 ---CC-----GLRLKGTDVVFTKR 483
D 468 SKFCFMKMLGQECCTCFLPABGAVDQGHNDNAYEGSDVDPABSAISLSSGVVPGTA 527
QY 484 VVDKRVHSL-----GDIICDVRL-----SP 504
D 528 LQPLVQALDLPAEIVARAGRLTATVAKVQVDGRIDCEITLIGNTPTPTS FVDAVLETNCP 587
QY 505 EQ-----VGPLRSVP----- 515
D 588 ERNLSFSDASQSTMAAGPSSLTYAASAGLEVRVYAAGLDHRVAPVAGVSPRASPGEVTA 647
QY 516 -----PAVFRHD----- 522
D 648 FCSALYRFRRAQGRHSLIGNLWHPREGILGLFAPRFPGHVWBSANPCCGSESTLYTRTWS 707

```

```

QY 523 -----RELEVLRENG-----CYNERPVSTPVEBPQ----- 551
Db 708 VDVAVSPARPDGFWSEPSIPSRATPTLAAPLEPPADPSPBPAPALASPAGTAGA 767
QY 552 -----PDADLMHAT----- 560
Db 768 PAITHQTAHRRLFTYPDGSKVPAGSLPESTCTMLVNASVNDHRPGGLCHAFYQRYPA 827
QY 561 ---AASL-----PEYR-----ATLOAG-----LN 576
Db 828 SFDAASFVARDGAAAYLTLPRIIHAVADYRLHNPKLEAAAYETCSRLGTAYPPLG 887
QY 577 TDVQO-----LKIT 585
Db 888 TGIYQVPIGSPFDAMERNRPGDELYLELAARMEANRPTRPILTTEDVARTANLIE 947
QY 586 LENAL---KTIDGLTSPVRLGEMYE---GPPSGGTGLTLLAALBAAGKALVYAPREL 639
Db 948 LDSATDVGACAGCAYTP--GVYQYQFTAGVPGSKSSTQTA-----DUDVYVPTREL 1000
QY 640 REAMDRRIKPPSASATQVALAIL--RRATAEGAPFATVVIDECMFPLVYVAIVHALSP 697
Db 1001 RNAMRRR---GFAAFTPHTAARVTQGR--VVIDEAPSLP--PHLLLH--NQR 1045
QY 698 SSRIVLVGDVHIOGIFIDFQGSANMPLVADVVKQCRRTFNQTKCPADVYATTFQSLY 757
Db 1046 AATVHLLCDPNOIPAIIDEHAGL--VPAIRPDIGP--TSMVHTHMPADV--CELIIRGAY 1100
QY 758 PGCTTSGCVASISHVADPYRNSAQOTCTOEERKRGAGMAGMVEHAEQRTFASVLH 817
Db 1101 PMIOCTTSKVLNSLFGEP---AVGOKLVFTQAAKPA--PGSVVHEHQATYETETII 1154
QY 818 YNGSTAEOKLLAEKS--HLVGTIRHTNHLVIRDPGTDIRQLNHSARAEVFTDIPAPLE 875
Db 1155 ---ATADARGLIOSSRAHAIVALTRHTEKCVIIDAAP--LIREVIGIS--DAIVNFPILAGE 1209
QY 876 ITTVPESEVQRN-----EWATIPPOASTPHGAIHLHARKNGDPDCCVALATGVEV 930
Db 1210 IGHOPSVYIPRGNPDANVDTLAAFPSCQI--SAFHQLAEBELGHRP--VVAVALPPECBEL 1266
QY 931 FGGRKINVELAEPDAPKPRHAPQ--EGVQWYKVTNASKHOALQTLISRYKRSADLP 989
Db 1267 EOGLLYLPOELTTCSSV---TFELTDIVHCRMAAPSKRAVLSTLVGRYGRRT---KL 1319
QY 990 HEAK--EDVKRMILNSLDHRMDWTVEDARDAVFETOLKFTORGTEVDELLSPD--DPYIR 1046
Db 1320 YNASHSDVR---DSLARFIPIAGVQVTTCELYELVEAMVEKQOGSANTLBDLCNRDVS 1376
QY 1047 DIDFLMKTOOK--VSEKPIINTGVGOGIAHKSLSLNFVLAAMIRILEB--ILRTGSTRYXS 1104
Db 1377 RTTFQKCKNKTETJETIAGGVGOGISAMSKTFCALFGMPFRAIEKILALLPGSVFVG 1436
QY 1105 NGLPREE--EAMMLEKINOVPHATFVSADWTEPDTAHHNTSELFPALLERIGTPAAAVN 1163
Db 1437 DAFEDTVSAVAALAAKASNV-----PNDSEFSDTQNNFSLGLSCALIEBGMQMLIR 1491
QY 1164 LFERECGRKTLRAKGLGSEVVDGL-----LDSGAATPCRNITFSAANVLT--FRGYKF 1216
Db 1492 LY-----HLIRSAMILQAPKESLGRFWMKSGEPTLLMNTVMMAVITHCYDRDFGV 1545
QY 1217 AAFKDDSL--LCSHYLRFDASRLHNGERYKTKHLKVEYQKIVPIYIGLVSAEQVYL--DP 1274
Db 1546 AAFKDDSLVLCSEYROSPPGAVALTAGCLK---LKVDPRPIGLYAGVVAVAGLALPDV 1602
QY 1275 VASA 1278
Db 1603 VARFA 1606

```

RESULT 5
US-10-223-070-19
; Sequence 19, Application US/10223070
; Publication No. US20030109045A1

```

; GENERAL INFORMATION:
; APPLICANT: NELSON, RICHARD S.
; APPLICANT: DING, XIN SHUN
; TITLE OF INVENTION: RNA SILENCING SUPPRESSION
; FILE REFERENCE: NBL:00605
; CURRENT APPLICATION NUMBER: US/10/223,070
; CURRENT FILING DATE: 2002-08-16
; PRIOR APPLICATION NUMBER: 60/313,185
; PRIOR FILING DATE: 2002-08-17
; NUMBER OF SEQ ID NOS: 24
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 19
; LENGTH: 1694
; TYPE: PRT
; ORGANISM: Raspberry bushy dwarf virus
US-10-223-070-19

Query Match      3.3%; Score 294; DB 15; Length 1694,
Best Local Similarity 18.4%; Pred. No. 5.8e-14;
Matches 306; Conservative 228; Mismatches 589; Indels 540; Gaps 76;

QY 47 YPLRFKGGTLPTQHPILAGHQRY--ABEVLNPFARGSTVLEIGPSLSALKHAGAPAP 105
Db 191 YELKFTQNVQGP--HNMAAARLLEJTHDLNSFP--ADAPILDIGNMFSHR--YGRSNV- 245
QY 106 VADYHGTCTKGTGDSRH---TLAE-----SRSVATGRPEF- 139
Db 246 ---HSCCPMLDIRNERHTRLTWTESIMSLRHRVAGTIDLPDAMLSTRVDSMKSEFY 302
QY 140 ---KADASLLANGTA--SRTPCVQGV-----SCAFKSRV 169
Db 303 KRAVHPDLRLVLAGIRDGNBSLYCHKFKVSNMDVMBEERNCLTVPEPECFKAKY 362
QY 170 GIANHSLDYTLLEBIANAENHGLMVAFAFMPEBELLYMDNVNABIGYEFHYIEBPMA 229
Db 363 AIAVHSGYDLPKLELIGMVGHVGLGWTIADPAVLV-----ATSGY----- 406
QY 230 VKDCAFQGGDLRLHPBELDINSQERRIERLARGSYRAVAFVPGDDMDGAYLHDFH 289
Db 407 ---IPALRCNWEKSKQIWFSPRDSITMGYRHDWE 438
QY 290 TWLAVLYRNPPTPFGSLHIEYQR---RHGSSIBELRTAPRQDRLAVVPRTSGCLC 345
Db 439 VYSKTLT---STVVSCKKFTYMERDKTRRGV-----LPSIINKSGSLR 480
QY 346 RIPNIFYY-----ADASGTE-----HKTILTSQHKYNNL----- 374
Db 481 KGDHTFFFNAMFHEMYDKYINKVPLVYKQDLYGDBGSVSCMRBYVMSRKLVDRVIEVCL 540
QY 375 ---LNFQTRPEKELVDKTLMSFAPARLRAIVASVETBSSWNISPADLVRTVSLY 429
Db 541 RGVPKINFGNCDDAVHMDLRIQNLHLSHQTLVANGSTIIRERAIPIKQFSPVSVIY 600
QY 430 VLIHIERRAVAVK---TAKDVRGETSPWES-----LKHVIGSCC 468
Db 601 FEILITRKESLSLAWFAGJAPDFKGS---WSSSLKRVYTRILGPANILKTYLNLFP 657
QY 469 GLRNKGTDVVFTKAVVDKRVHSLGDIICDVRSLSPQVQ--FLSPRVPAEVFHDREBLE 527
Db 658 RCRD--KVSDEMFPVPAVAKLVV-----LENTYIGSLMGDDCTLEKYDSSAFN 705
QY 528 VLREAG--CYNERPV---PSTPP----- 545
Db 706 ILENVGNELFNNSSTDSGKPEPTVMTGNPNAVIAEASICYRAVRDIGKCKERICLHAY 765
QY 546 ---VERPOGFDDL--WHATA-----ASLPEV----- 567
Db 766 QATNGCGYLANDTVNGVFDKMTSVQKPKFDEHFGDSSFTYLSFGKIPDPVGYL 825
QY 568 ---RATLOAGINTDVQKQKITLENALKTIDGLTSLPVNGLBMYEGPSPSGKGTGLIA 621
Db 826 VVTOGTRVTTMLKFSROVATIPATVTPITIKVDDGT-----GQKKTIELVR 871

```

```

QY 622 ALBAAGKALVYATRELBREAMDRIKPSASATOHVALILRRATAGAPATVTVDEC 681
DB 872 RYRFG---LILISVCANADEIRKLAANDSKIRTYDSL-----SPSYTGCDEL 921
QY 682 FM-----PFLVVAIVHALSPSSRIYLVGDVHQIGFIDQTSANMPL-----VRDVVK 721
DB 922 FIDVYGLSHGILLALH-1SGIRKVTLPFDSQDIPPCN---RLADPPLKYNSEDVGLN 977
QY 732 CRRRTNOKRCADAVVATTFQSLY--PQCTTSGCVASIS-----HVAADYRNS 780
DB 978 FDEIRSTTYRCPODI--TLSLQKMYTKPIKTVSTESSITTKPIKSEBEIDLPJNAFQD 1035
QY 781 QAQTLFOTEEKS-----RHGABG---AMTVHAQGRTPASVILHYNAGSAEQKLAEXS 832
DB 1036 PULYICHTKHDESLLKRWAKENISSEVRYVHAAGLSYKQNV-YERLITDNDLTYTKRK 1094
QY 833 --HLVGTIRHTNHLIYRDPDGIEROLNHSKAEVTTDIPALEITTVPSSEV----- 885
DB 1095 LPVHLVAISHHTDKI-----VYCTTKPESSDPSLSALKNTITKSRDLQGEAS 1142
QY 886 -----QNEVMATTPPOSATPHGAIHLRKXFGDQPCGCVALAKTGEVFG-- 932
DB 1143 GSESSYAVVFNESNEVATKPEVCENVRKAEM--NFPVSD-----ALYQKEVPIYGA 1195
QY 933 ---GRAXIN-----VELAEPDATPKPHAFQ-----GVQW--VKVTN-- 965
DB 1196 PDKGKASINPGSVIRAIIEIIPGNTSIDTALDELVEVPMSLQVGSIMDVSKISPL 1255
QY 966 -ASNK---HOALQTLISRYT-----KRSADLPHEAKEDVYKMLN-SLDRHMD 1009
DB 1256 FTNNKFAVPHLPFGALLRRTTSROVGLAIKENANVMNSQKYPDLLENLANKAVRFPD 1315
QY 1010 TVTEDARDR-----AVETOLKFTQGTVED--LLEPDPY---INDIDFLMK 1053
DB 1316 FIDMEKFSKLPYGLVSSAEQIOTYQNKTKNKTDPYCVALSPIQYKXMKIKEDVKEFLT 1375
QY 1054 -----TQOKVSPKPIINTGVGOG-----IAHSKSLNFWLAAMIRILBEILRT 1096
DB 1376 DGAQSEYTKAATTYHQPEITQVATALFQPKTRLLACRNKFLNIP----- 1422
QY 1097 GSRTRVYSNGLPDEBEAMLEAKINQVPHATFVSADTEPDTHANNNTSELLFALLERIG 1156
DB 1423 -----EHNDLSEVLYTKYHGSENN-----TFTEIDSKFDKSGEIHQILQDILILKFG 1472
QY 1157 TPAAVNLFRERCCKRTLAKGLG-SVEVDGLDLSGAAMTPCRRNTJFSAVMTLF-- 1211
DB 1473 CDEPVALMSTARRSSIFQNVGIGFKTDFORRTGDAFTFLGSLVTAAMLAFAVISDP 1532
QY 1212 -RGVKAAPFKGDSLSCGSHYLAFDASRLMGERYKTKHLKVENYQKIYPIYG--LLVSA 1267
DB 1533 REKIRYMLVGGDSLCS--YGPLOVLEPLGDI FNMS-CKL-VOPACPYFASRYLIRG 1588
QY 1268 EGV--VLDPVRSALKITFGRCYTSSELYSKYVEAVRDTTKGMSDARYSLCHMSACYNY 1325
DB 1589 DELICVPDPYKLVKL-GR-----KQVPDN-----QASICEIR---TG 1622
QY 1326 APESAAYIIDA VVRFGGDFPFEOQLRV--VRAHQAOPDAYSS 1365
DB 1623 LADSAKYIFDDIVK-----QKLALVQVRYNKAPSLYDA 1657

```

```

RESULT 6
US-10-224-999A-3475
; Sequence 3475, Application US/10224999A
; Publication No. US20030171318A1
; GENERAL INFORMATION:
; APPLICANT: Myriad Genetics, Inc.
; APPLICANT: Morham, Scott
; APPLICANT: Zavitz, Kenton
; APPLICANT: Hobden, Adrian
; TITLE OF INVENTION: Composition and Method for Treating Viral Infection
; FILE REFERENCE: 5004.01
; CURRENT APPLICATION NUMBER: US/10/224,999A

```

```

; CURRENT FILING DATE: 2003-03-03
; PRIOR APPLICATION NUMBER: US 60/313,695
; PRIOR FILING DATE: 2001-08-20
; NUMBER OF SEQ ID NOS: 3484
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3475
; LENGTH: 2116
; TYPE: PR
; ORGANISM: Rubella virus
US-10-224-999A-3475

Query Match
Best Local Similarity 2.9%; Score 253.5; DB 12; Length 2116;
Matches 187; Conservative 124; Mismatches 347; Indels 223; Gaps 41;

QY 600 PVNG---LEMYRCPGSGKGTLLIALBAAGKALVYATRELBREAMDRIKPSASATQ 656
DB 1340 PRKGPYNIRVMNMAAGAKTTRILAAFT---EDLYVCPTNALHIEIQALBARDDIKN 1396
QY 657 HVALALIRATAGAPATVVIDECMFPLVYVAYVHALSPSSRIYLVGDVHQIGFIDQ 716
DB 1397 --AATYERALTRELAARYIIDEAFTLGEYCAFY-ASQYTAEVICVGRDQCG----- 1448
QY 717 GTSANMPLVADVVKOCRRRTFNQTKCPADVATTFQSLYPCCTTSGCVASISHVAP 776
DB 1449 -----PHYANCRTPV--PDRWPTERSRHTV--RFPDCW-----AARLPAQLD 1487
QY 777 YRNSAQOT-----LCFTQEKSRHAGB--AMTVHAQGRTPASVILAH 818
DB 1488 YDIGERTGFACNLMDGRQVDHLAFSRETVARLHAGIRATVHAQMSVGTACTIH 1547
QY 819 NGSTAQKLAESHLVGTIRHTNHLIYRD-----PTGDIROLNHSKAEVTTDIPA 872
DB 1548 GRDGTVALVALVDLAIVSLTRASDALYLHELEDGSLRAAGLSAFLDAGLAEI-KEVPA 1606
QY 873 PLE-ITTVK-----PSEVQNEVMATIP-----QSATPHGA----- 904
DB 1607 GIDRVVAVEQAPPLPDPADIPBAQ--DVPKCPPTLSELVYGRGHFHYADLNRTGEG 1663
QY 905 -----HILRKXFGDQPCGCVALAKTGEVFGGRAXINVELAEPDATPKPHAF 954
DB 1664 EREVRVYRISRLNLNHTNTEPTEVLSAV-----CAVRRIRAG 1703
QY 955 QSEGVQVYKVTNASKIOALQTLISRYTKSADLPHEAKEDVYKMLNSLDRHMDWTED 1014
DB 1704 EDG-----STLTAVAROHPR--PFIQIP--PRVTAGVAGWRMTYLRE 1744
QY 1015 ARDAVETOLKFTQGTVEDLLEPDDPYIR--DIDPLMTQOVKVSPP----- 1061
DB 1745 RIDLTVYTTQGVAAAR-----ELTDYARRYPETIPAGMCTAQSLSVAPLKAATLKCV 1796
QY 1062 -----PINT-----GKYQGGIAHSKSLNFWLAAMIRILBEILRTGSR--TVRYSNGLP 1108
DB 1797 DALGPRDTEDEGAAGKGLERAKAKENVQMSHPFALQITIMBALPQFLVAAGHT 1856
QY 1109 DEEBAMLEAKINQVPHAT--FVSADWTEFD--TAHNTSSELLFALLERIGTPAAAVN 1163
DB 1857 EPEVDAMWQA-----HYTNAIEVDFTBEDMNOQTALTRDVELEISAL--LGLPCAB-- 1906
QY 1164 LFRERGCKTLRA-----KGLGSEVVDGLDLSGAAMTPCRRNTJFSAVMTLF--RGVKE 1216
DB 1907 -----DYRLRAGSYCTLEBELSTTEGCRSTGSPATLLANTTVAAQMMARRVPKQVR 1960
QY 1217 AA-FKGDSSL-----CGSHYLFDAASRLM--GERYTKTLKTVENQKIYVYIGLVSAG 1269
DB 1961 AGIFQGDWVITFLPEGARSAALKMTABAVGLPFGPHIPVKHVSPTPBGCHVG---TAAG 2017
QY 1270 VLDPVRSALKITGRCYTSSELYSKYVEAVRDTTKGMSDARYSLCHMSACYNTYAPES 1329
DB 2018 LFDVHQAOKVLCRRFPDPLVEEQOV--ALLDLRLQGYVALPPTVAA--NAAAYDYSAER 2074
QY 1330 AAYIIDA VVRFGG---DPP-----FEOQLRVAVAHQAOPDA 1362

```

DB 2075 VLAIRELTAARAGGLDHPATIGALBEIQTYPAPANLHDA 2115

RESULT 7
US-10-223-070-4
Sequence 4, Application US/10223070
Publication No. US20030109045A1
GENERAL INFORMATION:
APPLICANT: NELSON, RICHARD S.
APPLICANT: DING, XIN SHUN
TITLE OF INVENTION: RNA SILENCING SUPPRESSION
FILE REFERENCE: NILE:00605
CURRENT APPLICATION NUMBER: US/10/223,070
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: 60/313,185
PRIOR FILING DATE: 2002-08-17
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 4
LENGTH: 961
TYPE: PRT
ORGANISM: Brome mosaic virus
US-10-223-070-4

Query Match 2.6%; Score 228.5; DB 15; Length 961;
Best Local Similarity 19.1%; Pred. No. 5.1e-09;
Matches 203; Conservative 151; Mismatches 330; Indels 381; Gaps 51;

QY 13 AAADVAVANVQORAVKUDFAPPLKALETLHL-----YYLRFKGG--TLPTQHP- 62
DB 20 SHODIVDNQVQAQSLAQLIYAKRSKINVRNKLSEBDAFDRDYGAFDLNLTOQYHAP 79
QY 63 -ILACHQVAAE-EVLNHFARGRSTVLEIGPILSALKLHGAPNAPVADYHG----- 111
DB 80 HSLAALALVAEHYDGLDSF-----PPEPVIDFGSGWMHHSR 117
QY 112 -----CTKYGTDDGSRHITAL-ESRSVATGRPEPKADASLANGLASRFFCVDGVS 162
DB 118 RDKRVHSCCPVLGVDRDAARHEERCMRKRKIOESDDFDEVPN-----FCINRAD 167
QY 163 CAFKSRVGIANHSLVDTLLEELANA PENNGLMVNAFMMPELTY----- 208
DB 168 CDVQADMAICIHGGYDMGFQGLCDMNHSHGVNLGTVMFGAMLPDREGPLPLKCHQ 227
QY 209 -----MDNVVN-----AELGYRF-----HYIEEPM----- 228
DB 228 RDGSGADEVIKDFENESTLSYIHGMODGSFTESVHICIDGTLYLEEMLKCNIMTYK 287
QY 229 -----AVKQCA-FQGGDLR--LHPELDFTN-----ESQERRIERLAK 264
DB 288 IIAITNLRCPRETILRHCVWFE-DISKYGVGSIPEDMSLRMKCVAKTVEVEIEIAR 345
QY 265 GSYSRRAVIFSGDDMM-----GDAY-----LHDTHTMLAYLL 296
DB 346 -----CFKESKEMTENKAVASILAKSSTVIINGQAIMAGERLDETHL-VAFML 396
QY 297 VNNYETPF-----GFSLHIEVQRRHGSSEILRITRAPGDMLA VVP--RTS 341
DB 397 TLNLQOKEXKLALNDGMEWMKGMCHFKTRPMMG-----GDSSRAKVGMLRTL 444
QY 342 QGLCRIPNIFYADA-----SGTEH-----KTILTSQHKMMLANFMQTRPE 383
DB 445 ASRPFLLRLDSYADSFKFLTRLISNVEEPQDSVPISRLRTFTWBEEDLPRLHEBVOJTAKT 504
QY 384 KELVMTVLSFARALRA-IYVASEVTSSMNISSP-----ADLVRTVSVLYLHLII 434
DB 505 K-----RSKKKAKVPPAPALTPQEFHDAPESSSPESVDDVAPVD-----V 547
QY 435 EERRAAVAVKADVDVFGETSFMESLKHVL-----SGCCGLRNL-----KGTDVVPTKEV 484
DB 548 PAAEVSVETPDPRGISRHGAKKEFRYCKRLHNNSESULRHLMDSISGRGSEIA-NKSI 606
QY 485 VDKRVHSLGDIICDVRLS-----PEQ-----VGLPFSVPPAPVPHDBEIEVLEACGC 534

DB 607 FETV--NRHIDMV-NVHLANGNNMLYPKKYDTVTGNEHGLBPK--HADERITYDKTCAC 660
QY 535 YNERPVPSTPVEBEQGFADLMHTATAALPEYRATLQAGLNTVDKQKITLENALKTID 594
DB 661 SNLR-----DIAEBSA-----KVSV----- 675
QY 595 GLTISPVGLMEYBPPGSGKTGTLLIALBAAGKALVYAPTRBELREMBRRRIKPSASA 654
DB 676 -----PTCDISMVGVACGKTITAIKDAFRM--GEBDLVTANRKSABDVPMALFPDITYNS 728
QY 655 TQVVALILBRATA--EGAPFA--TVVIDECMPFLVYVALVHALSPSSRIVLVDVNOI 710
DB 729 --KVALDVARTADSALMHEGVSCHRLVDKGLLHYGOLLVAAALSKCSQVLAFGDTQO- 785
QY 711 GFIDFGTSANMPLVDVVKOCRRRTFNQTRCPADVATTFP-----QSLYPGCTT 762
DB 786 --ISFKSRDAPFKLHGNLQYDRDGVAKTRCPQDVIAAVNLKRGKGNRDTKQGSWTS 843
QY 763 -----TSGCVASISHVADYRNSQAQTLCTQSEK-----SRHGA 797
DB 844 ESKVSRSLTKRRITSGLOVTI-----DNRRTVLTWTDKALQTRAKDPVSKWMI 895
QY 798 EGAM-TVHAGRFPASVILHNGSTAQKLAEKSHLVGITHR 841
DB 896 DGHITVHEAOGISVDNVTLVRLKST-KCDLPKHEBYCLVALTRH 939

RESULT 8
US-10-388-848-2
Sequence 2, Application US/10388848
Publication No. US20030182684A1
GENERAL INFORMATION:
APPLICANT: Dinesh-Kumar et al.
TITLE OF INVENTION: TOBACCO RATTLE VIRUS VECTORS AND RELATED COMPOSITIONS AND METHODS
FILE REFERENCE: YU-P01-014
CURRENT APPLICATION NUMBER: US/10/388,848
CURRENT FILING DATE: 2003-03-14
PRIOR APPLICATION NUMBER: 60/364,901
PRIOR FILING DATE: 2002-03-14
NUMBER OF SEQ ID NOS: 40
SOFTWARE: PatentIn version 3.2
SEQ ID NO 2
LENGTH: 1707
TYPE: PRT
ORGANISM: Tobacco rattle virus
US-10-388-848-2

Query Match 2.4%; Score 213.5; DB 12; Length 1707;
Best Local Similarity 17.8%; Pred. No. 2.3e-07;
Matches 300; Conservative 210; Mismatches 570; Indels 601; Gaps 67;

QY 65 AGHORVAEVLNHFARGSTVLEIGPILSALKLHGAPNAPVAD--YHGCTKYGTRDGR 122
DB 168 SKRQMSNEBELMTYQFRENI-----AARPAVACNNTYQGCTCRGSGDKK 213
QY 123 HTTALERSVANGREDFRADASLANGLASRTFCVDGVGSCAFKSRVGIANHSLVDTLE 182
DB 214 K-----GAQYALAHSLVDYFLK 231
QY 183 EL-ANAFERNGLMTRATMAMPBELLYMDNVYNAELGYRFAVIEBPMAVKCAFGGDLR 241
DB 232 DLMATVBEKTKVYVAAALFAPESNL-----VDBGFL----- 263
QY 242 LHPELDFTNLSQERRIERLAAAGSYSR--AVIFSGDDMDGDAYLHDTHTMLAYLVN 299
DB 264 ---ESVD-----GYMKKNGKITGFEDSPSIIHDEBKYKTYLGLK- 303
QY 300 YETPF-GFSLHIEVQRRHGSSEILRITRAPGDMLA VVPRTS-----QG 343
DB 304 -PVSQGVAVFYEPWQVGDWTLFSIYR-----IAGVPRSLSSQBYRARIYSRMEN 355
QY 344 LCRIPNIFYADASGTEHTKITLSQHKVNMMLNFMQTRPEKELVDMTVLSFARALRAI 403

```

Db      356  M V V P I F D L V E S T R E L V K D L F V E K O P M D C L D Y I A R L S D Q L T - I S N V S I S S N N W V L      414
Qy      404  V V A S E T S E S S W I S P A D L V T V S L Y V L I I E R R R A A V A V K A D V E G S T S W E S H K V      463
Db      415  F I N G A A V K N K O S V D S D L Q L A Q T L - L V K E Q V A P V R E I R E A L I T E T K D I T S L D V      470
Qy      464  I G - - - - - S C G L R L K G D V V F - - - - - T K R V      484
Db      471  L G I S R K L M K O P A N K I A V G F V G A V G T L I G F Y P K V L T W A K D T P N G E L C Y E N S H K T K V I      530
Qy      485  V D K Y R V H S L G D I - - - - - I C D - - - - -      499
Db      531  V F L S V V Y A I G C I L T M R D I R D G L V K K L C D M F D I K R G A H V L D V E N P C H Y E I N D F S S L Y S      590
Qy      500  - - - - - V L S P E Q V - - - - -      507
Db      591  A S E S G E T V L P D L S E V K A S D K L L Q K K E I A D E F L S A K F N S V G S S V T S P S V V G S S R S G      650
Qy      508  - G P L - - - - - P S R V P P A V P H D R E L E - - - - -      527
Db      651  L G L L E D S N V L T A R A V G S R K V D D E I M E Q F L S G L I D T E A I D E V V P A F A S A C E R G E T S G      710
Qy      528  - - - - - V L R E A G C Y N E R P - - - - - V P S T P P Y E B P Q G F P A D -      555
Db      711  T K V L C K P L T P R G E N V L P A V K P L V S K G K T V K R V D Y F O V M G G E R L P K R P V I S G D D S V D A R      770
Qy      556  - L M H A T A A S L P E - - - - - Y R A T L Q A G L N T - - - - - D V K O L K - - - - - T L E N A      589
Db      771  E F L Y V I D A E R V A Q N D E I M S L Y R D V S R G V I R T G G O N Y P H L G V D V E K M C I R P V T E H A      830
Qy      590  - - - - - - - - - - - - - - - - - - - - - - - - - - - L K T I D G L      596
Db      831  Y V F Q P D R M D W S G Y L E V A V W E R G M L V N D P A V E R M S D Y I V C D Q T Y L C N N R L I L D M L S A L      890
Qy      597  T L S P V R - G L E M Y G P P G S G K T G T L I A L B A A G K A L V - - - - - A P T R E L B E - - - - - A M D      644
Db      891  D L A P V A C S F E L V D G V P C G C S T M I V A N S A N C D V I T S G A A I D D L E R A S K G F P E K L K      950
Qy      645  R R I K P S A S A T O H A L I L R R A T A E G A P A T V V - I D E C M F P L V Y V A I V A L S P S S R I V L      703
Db      951  R A V K A T V D - S E L M H - - - - - C V D S L T G D V L H F D A L M A H A G M V F C A Q I A G A R K C I C      1000
Qy      704  V G D V H O I G F - - - - - I D F O C T S A M P L V R - - - - - D V K O C R R T E N O T K S C P A D V A T T F F O S      755
Db      1001  O G D O N O I S F K P R V S Q V D L R F S S - - - - - L V G K F I D V T E K R - - - - - E T Y R S P A D V A A V - L N K      1049
Qy      756  L Y P G C T T T S G C V A S I S H V A P D Y R N S Q - - - - - A Q T L C F T O E K S - - - - - R H G A E G A M      801
Db      1050  Y T G D V R T H N A T A N S M T V R K I V S K E O V S L P R G A Q Y I T F L O S E K K E L V N L A L R K V A K V S      1109
Qy      802  T V H A O G R T P A S Y I L H Y N G S T A E O K L A E K S H L V G I T R H T N H L - Y I R D P T G D I E R O L A N H      860
Db      1110  T V H E S G E F T K D V L V R T K P T - D D S I A R G R E Y I L V A L S R T I O S I V T V E D V S K I R E      1168
Qy      861  S A - - - - - K A E V F T D I P A P L E I T T V K P S E E V O R N E V M A T I P O S A P - - - - - H G A I - - - - - H L L      908
Db      1169  S A A L T Y A A L A R F V T E T V L Z R F R S R P D V F R H H E G P C A V P D S G I T T D L E M Y Y A L L F P G N S L      1228
Qy      909  R K R F G S G - - - - - Q D D C G V U L A K T G Y E V F G R A K I N V E L A P R A T P R K H A F O G S V G W K V Y      963
Db      1229  R O S S L D G Y L V A T T D C N - - - - - L R L D N V T T K S G N M K - D K F A K E T F L K P - - - - - V I R      1273
Qy      964  T N A S N K H O A - - - - - L O T L S R Y T K R S A D L P L H E - - - - - A K E D V K M L N S L D R H M D W T V E D      1014
Db      1274  T A M P D R K T T Q L E S I L A L O K R N Q A A P D L O G N V A T V L I E B T M K K L S V - - - - - Y D G K I      1328
Qy      1015  A R D P A V E T O L K - - - - - F T O R G G T V E D L E P D D P Y I R D I D - - - - - F L A K T O O K V S P K P I N T S K V G      1069
Db      1329  R A D P I V R A Q M E R M W R N Q S T A V O A K V A D V A R E I H E I D Y S S Y M W I K S D V K P K D L T P O F E      1388
Qy      1070  - - - - - O G I A H S K S L N F L A A M I L E E I L T G S R - - - - - T V R Y S N G L P R E E E A M L E A K      1119

```

```

Db      1389  Y S A L Q T V V Y H E K L I N S L F G P I P K E I N R K D A M Q P H P V F T R M T S S D L N D R V K F E L N T B A A      1448
Qy      1120  I N O V P A T P V S A D W T E P D T A N N T S E L F P A L L E R I G T P A A V N L F P E R C G K T L R A K G L      1179
Db      1449  Y D - - - - - F V E I D M S F D S A N F H L O L E I R L F G L D E M A A F L M E V S H T O T T V R - - - - -      1498
Qy      1180  G S V E V D G L L - - - - - D S G A A W T P C R N - - - - - T I P S A A V M L T L P F G V K P A A F K G D S L L C G S      1229
Db      1499  - D I Q - N G M A H I W Y O O K S G D A D T Y N A N S D R T L C A L L S E L P L E A V - M T T G G D S D L I A P P      1555
Qy      1230  H Y L R F - D A S R L M G E R Y K T K L K V E O K I V - P Y I G L V S A B O V L D P V R S A L K I - - - - -      1281
Db      1556  R G Q F V D P C P K A T K N N F E K I F K Y D V P M W C G K F L K T S C Y E R V P D P V A V L T L G K S K I      1615
Qy      1282  - - - - - F G C T S E L Y S K Y E A V R D - I T G - - - - - W S D A R Y H S L      1315
Db      1616  K D V O H L A E I Y I S I N D S R A L G N T M V V S K L S E S V S D R Y L G D S V H A L C A L M K H I K S P T A L      1675
Qy      1316  C      1316
Db      1676  C      1676

```

RESULT 9

```

US-10-138-842A-2
; Sequence 2, Application US/10138842A
; Publication No. US20030148390A1
; GENERAL INFORMATION:
; APPLICANT: GONSALVES, DENNIS
; APPLICANT: LING, KAI-SHU
; TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS PROTEINS AND
; FILE OF INVENTION: THEIR USES
; FILE REFERENCE: 07678/025006
; CURRENT APPLICATION NUMBER: US/10/138,842A
; PRIOR FILING DATE: 2002-05-03
; PRIOR APPLICATION NUMBER: US 09/579,259
; PRIOR FILING DATE: 2000-05-25
; PRIOR APPLICATION NUMBER: US 09/224,898
; PRIOR FILING DATE: 1998-12-31
; PRIOR APPLICATION NUMBER: US 08/770,544
; PRIOR FILING DATE: 1996-12-20
; PRIOR APPLICATION NUMBER: US 60/009,008
; PRIOR FILING DATE: 1995-12-21
; NUMBER OF SEQ ID NOS: 67
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1390
; TYPE: PR
; ORGANISM: Grapevine leafroll virus
US-10-138-842A-2

```

Query Match 2.4%; Score 213; DB 12; Length 1390;

Best Local Similarity 20.3%; Pred. No. 1.8e-07;

Matches 213; Conservative 118; Mismatches 326; Indels 394; Gaps 50;

```

Qy      60  Q H P I L A G H O R V A E S V L H N F A R G S T Y L E - - - - - I P S L H S A L K L H G A P N A P A D Y H G C T K Y G      116
Db      494  E H P - A C G E A E V E R D L I T P L G - - - - - T A V L E S P P V G P E A G S - - - - - A P V - - - - -      530
Qy      117  T R D G S H I T A L E S R S A T G R E P F K A D A S L L - - - - - A N G I - A S R T - - - - - F C V D G V S C A P      165
Db      531  V E D G C P E V E A K E S E V I V D V S S E P V Q E V L S T N G V Q A R T E B V V O G D T C G A V A K S E V      590
Qy      166  K S R V G I A N S L Y D V T L E - - - - - E L A N A F E N H G I A M V R A F M H      202
Db      591  S Q R V F P A Q V A H A G L E A S S A G A V E P L O V S V P A V E K T V L S V E K A E L K A V D K A V H A      650
Qy      203  P E E L I T M D N V A N A L G R F H V I E P A V K D - C A F O G G D L R L H F P E L D F I N E S O E R R I E R L      261
Db      651  K E - - - - - V K N V P V T L P R G A L K I S E D V R K E L C M F R I C S C G V - - - - - Q L D V N E A - - - - - T I      697
Qy      262  A A R G S Y S - - - - - R R A V I F S G D D W K G D A Y - - - - - L H D F H T M L A Y - - - - -      294

```

```

698 ATRFSNATFYDLSKGRSAVFPS---KLGEYTYNGSGHVSQMPRALEDILITAIKTPSV 754
QY 295 ---LLVRYPTPPGFSLSIEVORRGSSELR-----TRAPGDMALAVPPTS 341
DB 755 FDHCLVQYKXKGGVPPFAHDEBECYPSDNPILTVNLVKANFSTCRGKGMVI----- 809
QY 342 QGLCRIPNIFYADASGTEHKTILTSLQHKVNL-----LNFQTR----- 381
DB 810 ---NVASGDYFLMPCGFQ---THLSVNSIDEGRISLTFRATRRVGVGRMLQLAG 861
QY 382 -----PEKELVDMTVL-----NSF 395
DB 862 VSDKSPGVNPOQSGATRTITPKSGKALSBGSGREVKRSTYSIWCEQDYRKCEW 921
QY 396 AARALRAIVASEVTSSWNISPADLVTVVSLVLAH---IERRRA-----VAVXTA 446
DB 922 LRADNPVALKPGYTPMTFVVKAGTSDAVVEYLKLAIGRTYRALMARNAIAYTTA 981
QY 447 KDVFGETSFWESLK--HYLGSCCGLRLKGTDVVFTKRVVDKRVHSLGDIICDRLSP 504
DB 982 EGVLLKVPQVYESLPGFHVYS-----GTDLIF-----HSTQDL----- 1016
QY 505 EQVGFLLPSRVPARVFDHREBLEVLREAGCYNERVPSTPVEBPQGDADLMHATAASL 564
DB 1017 -RVNDLP-----YVF-----IAEKGIF-----IKGQVD----- 1039
QY 565 PEYATLOAGLNTDVYKQKITLNLKTIIDGILTSPVAGL-----EYEGPSSG 614
DB 1040 ---AVVALGDNLSVCDLILVFHDAINLMGALKVARC--GMVGESPFSEYKCYNAPPGG 1094
QY 615 KTGTLIALLEAAGKALVYAPTRBLRBMNDRIKPPSASAT-----QHVALILRR- 665
DB 1095 KTMILV-----DEVKSPNSTAITITAVNGSESDIMAAVKGD 1131
QY 666 ATAGCAPAT-----VVIDECFNPPLVYVAIVHALSPSSRIVLGVH 708
DB 1132 PNLGLNATTVNSRVNFIYRGWKYRLVDEVVMMQQLQLGVFATGASGLFFGDIN 1191
QY 709 QIGFLD-----FOGTSANMLVNDVYKQCRRTFNOKRC-PADY 747
DB 1192 QIPFINREKVRMDCAVEVPKKEVYVYTSKSYRCL--DVCYLLASMTVRGTEKCYPERV 1249
QY 748 VA---TTFEFSILYPCCTTSGCVASISHVAPDYRNSQAQTLCTOEKS--RHAGG--- 799
DB 1250 VSGDKPVPYRSLSKRPIGTDDVAEIN--ADYV-----LCMTLESMDKRSIKGK 1300
QY 800 ---AMTVEAQRTPASVILHYNGSTAEOKLIAEKSHLVGITRRTNHL----- 845
DB 1301 ETPVMTVEAOGKTSFVVL--FRTKKADDSLFTKQPHILVGLSRTSLVVAALSSELD 1359
QY 846 ---YIRD--PTGDIERQLNHSAKAEVFTDI 870
DB 1360 KVGTYISDASPOSVDALHTFAPAGCFRGI 1390

```

RESULT 10

US-10-039-112-2

Sequence 2, Application US/10039112

Publication No. US20030198942A1

GENERAL INFORMATION:

APPLICANT: Goncalves, Dennis

Inventor: Kai-Shu

TITLE OF INVENTION: GRAPEVINE LEAFROLL VIRUS

PROTEINS AND THEIR USES

NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP

STREET: Clinton Square, P.O. Box 1051

CITY: Rochester

COUNTY: New York

COUNTRY: U.S.A.

ZIP: 14603

COMPUTER READABLE FORM:

```

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/039,112
FILING DATE: 31-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60009008
FILING DATE: 21-DEC-1995
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 19603/621
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1390 amino acids
TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULAR TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-10-039-112-2
Query Match 2.4%; Score 213; DB 12; Length 1390;
Best Local Similarity 20.3%; Pred. No. 1.8e-07;
Matches 213; Conservative 118; Mismatches 326; Indels 394; Gaps 50;
60 QHPLAGHQRABEVLNHPARGSTVLE---IGSLHSALTLHAPNAPVADYHCTKYG 116
DB 494 EHP--ACGAEVEKELIPLPLG---TAVLESPPVGBAGS---APN----- 530
QY 117 TRDSRHTTALSRSVATGRPEFADASLT---ANGI--ASRT-----PCVDGSGCAF 165
DB 531 VEDGCPREBAKCEVIVDVSSSEPPQVQEVLESTNGVQAATBEVVOGDTGAGVAKSEV 590
QY 166 KSRVGIANHSLYDTLE-----ELANAFENHGLMVRAPFMM 202
DB 591 SORVPPAQVPAHBAGLEBASGAVVEPLQVSVPAVEKTVLSVBRKARLEKAVDKGAVVHA 650
QY 203 PBEILLMDNVNNAELGTFPHVIBEPMAVKD--CAEQGDLRLHPBLDPIINSQERRIRL 261
DB 651 KE---VKNVVPKTLPRGALKISDVYKXELCMFRTCGCV---QLDVYNEA-----TI 697
QY 262 AARGSVS-----RAVIFSGDDDWGDAY-----LHDPHTMLAY--- 294
DB 698 ATRFSNATFYDLSKGRSAVFPS---KLGEYTYNGSGHVSQMPRALEDILITAIKTPSV 754
QY 295 ---LLVRYPTPPGFSLSIEVORRGSSELR-----TRAPGDMALAVPPTS 341
DB 755 FDHCLVQYKXKGGVPPFAHDEBECYPSDNPILTVNLVKANFSTCRGKGMVI----- 809
QY 342 QGLCRIPNIFYADASGTEHKTILTSLQHKVNL-----LNFQTR----- 381
DB 810 ---NVASGDYFLMPCGFQ---THLSVNSIDEGRISLTFRATRRVGVGRMLQLAG 861
QY 382 -----PEKELVDMTVL-----NSF 395
DB 862 VSDKSPGVNPOQSGATRTITPKSGKALSBGSGREVKRSTYSIWCEQDYRKCEW 921
QY 396 AARALRAIVASEVTSSWNISPADLVTVVSLVLAH---IERRRA-----VAVXTA 446
DB 922 LRADNPVALKPGYTPMTFVVKAGTSDAVVEYLKLAIGRTYRALMARNAIAYTTA 981
QY 447 KDVFGETSFWESLK--HYLGSCCGLRLKGTDVVFTKRVVDKRVHSLGDIICDRLSP 504
DB 982 EGVLLKVPQVYESLPGFHVYS-----GTDLIF-----HSTQDL----- 1016
QY 505 EQVGFLLPSRVPARVFDHREBLEVLREAGCYNERVPSTPVEBPQGDADLMHATAASL 564

```



```

Db      1017 -RRRDP-----YVF-----TAEKGF-----TKGVD-----1039
Qy      565 PEYRATLQAGLNTDVOKITTEENALKTIDGLTLPVRGL-----EMTEPGSG 614
Db      1040 ----AVVALGDNLSVDDILVPHDALINLMGALVABC--GMVGSFKSPFKCNAPGGG 1094
Qy      615 KGTGLTAALEAAGKALVYAPTELEBRANDRIKPPSASAT-----QVALAILRR- 665
Db      1095 KTMVLV-----DEPKSPSTATITANVGSSSDIMMAVKRD 1131
Qy      666 ATAEAGPAT-----VVIDECFMPFLVYVAIVHALSPSRIVLVGDVH 708
Db      1132 PNLGNSATTVNSRVNPIVRGMYKRVLDDEVYMMHOGLLQGVPAFGASBGLPFGDIN 1191
Qy      709 QIGFID-----FOGTSAMPPLVRDVYKOCRRTEFNOTKRC--PADV 747
Db      1192 QIPFINREKVFEMDCAVFVPKESVYVTSKSYRCPL--DYCYLLSMTVGTCKCYPEKY 1249
Qy      748 VA--TFPOSLYPGCTTSSGCVASISHVAPDYRNSOQTLCTQBEKS--RHGAG-- 799
Db      1250 VSGKDPVVRSLSKRPGLGTTDVAEIN--ADV------LCMTOLEKSDMKSLAGKGR 1300
Qy      800 ----AMVHEAGERTFASVILHNGSTAECKLAEKSHLVGIRHTMHL----- 845
Db      1301 ETPVMTVHEAQGKTFSDDVL--FRTKADDSLFTKQPHILVGLSRHTSLVYALSELDD 1359
Qy      846 ----YIRD--PTGDIROLNHSAKAEVFTDI 870
Db      1360 KVGTIISDASPOSVDALHTFAPAGCFROI 1390

```

RESULT 11

```

US-09-769-787-109
; Sequence 109, Application US/09769787
; Publication No. US20030091577A1
; GENERAL INFORMATION:
; APPLICANT: Microbial Technics Limited
; APPLICANT: Gilbert, Christophe FG
; APPLICANT: Hansbro, Philip M
; TITLE OF INVENTION: Proteins
; FILE REFERENCE: PWC/B21129MC
; CURRENT APPLICATION NUMBER: US/09/769,787
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: GB 9816337.1
; PRIOR FILING DATE: 1998-03-27
; PRIOR APPLICATION NUMBER: US 60/125164
; PRIOR FILING DATE: 1999-03-19
; NUMBER OF SEQ ID NOS: 388
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 109
; LENGTH: 1236
; TYPE: PRF
; ORGANISM: Streptococcus pneumoniae
US-09-769-787-109

```

Query Match: 2.3%; Score 205; DB 11; Length 1236;

Best Local Similarity 19.4%; Pred. No. 6,7e-07;

Matches 181; Mismatches 128; Indels 222; Gaps 28;

```

Qy      851 TGIEROLNHS-----AKAEVFTDIPAPLEITTVKSEEVQRN 888
Db      2 TETVEDKVSHTIGLDILKGIVAAGAVISGTVAOTQVFTNESAVLE-KTVEKTDALATN 60
Qy      889 E--VMAITTPQASATPHGAHILKRNFGDDPCGCVALAKGYEVFGRAKINVELAPDA 946
Db      61 DTVVLGTT-----STLSA-----SSTLSASEASSTASASASTSA 97
Qy      947 TPRHRAFOEGVGVVKTNASNKHQALQTLISRYTKSADLPHEAKED----- 995
Db      98 STASSTASASAS-----TSASTSISASTSVGSGTAATEATKAKVEDRKKPASYVAS 153
Qy      996 -----VKMLNLSIDRHWDMVTVEDARAVFETOLKFTQNGGVTEDLLEPDDP 1043

```

```

Db      154 VTMVNLQSAKRRKRSVDSIEQ-----LASTKMAAVFS-----GNTIVN-----GAP 136
Qy      1044 YIRIDDFLMKTOQKQSPKPIINTGKVGQGIAAHSKSLNPLYLAAMIRILEIARTGSR---- 1099
Db      197 AINASLIMAKSETKV-----YTGEV-----DSYVRVPIYKMLKVTNDGSKLTF 241
Qy      1100 -TYRYSNGLPDEBEAMLLEAKINQVPHATFVSADMTREPTAHNTEBELLPALLEITGP 1158
Db      242 YVTYVNA-----PKTNLDLN-----ISSMRGYSIYNSGISTQMTLTGSDLGK 286
Qy      1159 AAAVNLFRRCGR-----TLRAKGLSVETDGLDLSG-----AAWT--PCBNT 1200
Db      287 SGVKNYITTDKNGKQVLSYNTSTWTTGSSGYTWGNGQMGFPKAKKGGLTSMTVPIYTG 346
Qy      1201 IFSAAVMTLPRGVKFAKGGDSLLCGSHYLRFDASRLHMGERY-----KTGHLKVEVO 1255
Db      347 -----DTSFTFTPYAARTD-----RIGINYNGGKVVESSTTSG 381
Qy      1256 KIVPYIGLVSAQVVLDPVRSAKIFGRCTSELLYKTYEAVRDITKGMDPARYHSL 1315
Db      382 SLGSKSLSVASQ-----SASASASTSASASASTSASASAS 418
Qy      1316 CHWSACYNYPAPESAAYITIDAVYRFGKDPFPFROQLRVVAHQAPDAYSSTYFANVRASC 1375
Db      419 TSASASASTSASASASTSASASASTSAS-----ASASTSASEASTSASASASTSA 469
Qy      1376 LDHVFPPROAAAPAGVATCAKPEPSSLTAKAGSATTSHVATGTAPESPMDAPANS 1435
Db      470 SASASTSASASA-----STASASEASTSASASASTSASASASTSASASASTS 524
Qy      1436 FSELTPETPSTSSSPSSSSDSSTCGNSLSCGDTARTTEDINSKPPQDRQSRSEC 1495
Db      525 ASGSASTSTASASTSASASTSASASTSASASASISASBSASTSASEASTSTASASTSASES 584
Qy      1496 LDRSGERTGS-SLTAPASPSPSPSEBARILANGPVAAATSPASAPSCATQDVARTTP 1554
Db      585 ASTSASASTSASASTSASASTSASASTSASASTSASASTSASASTSASASTSAS 644
Qy      1555 DPAPFLGOSARAVSKPYRPTTARKKEVYPLAMKGVGTGDRPREVEDPETAIVVOLIS 1614
Db      645 ASASTSASASA--STASASASTSASASTSASASTSASASTSAS-----ESASTSASASTS 698
Qy      1615 GRYPQTKLSSDASKGYSTKGCSTGSPAPASADYQARDCCQTVRCRAAAEARSCTIHE 1674
Db      699 ASASASTSASASTSASASTSASASTSASASTSASESASTSASASTSASASTSASASTSASGS 758
Qy      1675 PLASSAASADLKRIKIRSTDSV---PDYKISKSA 1704
Db      759 ASTSTASASTSASASTSASASTSASASTSASESA 791

```

RESULT 12

```

US-09-851-410-2
; Sequence 2, Application US/09851410
; Publication No. US20030124510A1
; GENERAL INFORMATION:
; APPLICANT: Reyes, Gregory R
; APPLICANT: Bradley, Patricia O
; APPLICANT: Bradley, Daniel W
; APPLICANT: Kravtzyanski, Krzyzstof Z
; APPLICANT: Tam, Albert
; APPLICANT: Fry, Kirk B

```

TITLE OF INVENTION: DNA sequences of Enterically Transmitted

NUMBER OF SEQUENCES: 20

CORRESPONDENCE ADDRESS:

ADDRESSES: Dehlinger & Associates

STREET: 350 Cambridge Avenue, Suite 250

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/851,410
FILING DATE: 07-May-2001
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 09/128,275
FILING DATE: <unknown>
APPLICATION NUMBER: US 07/681,078
FILING DATE: 05-APR-1991
APPLICATION NUMBER: US 07/505,888
FILING DATE: 05-APR-1990
APPLICATION NUMBER: US 07/420,921
FILING DATE: 13-OCT-1989
APPLICATION NUMBER: US 07/367,486
FILING DATE: 16-JUN-1989
APPLICATION NUMBER: US 07/336,672
FILING DATE: 11-APR-1989
APPLICATION NUMBER: US 07/208,997
FILING DATE: 17-JUN-1988
ATTORNEY/AGENT INFORMATION:
NAME: Petibhory, Joanne R.
REGISTRATION NUMBER: 42,995
REFERENCE/DOCKET NUMBER: 4600-0183.24
TELECOMMUNICATION INFORMATION:
TELEPHONE: (650) 324-0880
TELEFAX: (650) 324-0960
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 431 amino acids
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-851-410-2

Query Match 2.2%; Score 198; DB 11; Length 431;
Best Local Similarity 26.2%; Pred. No. 4e-07;
Matches 106; Conservative 58; Mismatches 172; Indels 68; Gaps 17;
QY 890 VMATTPQSATPHGAIHLKRNFGDQPPGCAALAKTGVGVGGAQKINVELAEPDAPK 949
DB 5 VAAVLPCEPELEQGLLYLPQ---ELTTCDSV-----VFELTD----- 39
QY 950 PHRAFOEGVQWVKNVNASNKHQALQTLRSYTKRSADLPLEAK--EDVYKMLNSLDRWD 1008
DB 40 -----IVHCRMAAPSGRAVLSLVGRYGGRT---KLVAHSHSDVR---DSLARFIP 85
QY 1009 WTVTEADRAVAFETOLKFTORGTVEDLEBD--DPYIRIDFLMKTQOK--VSPRPINT 1065
DB 86 AIGPVQVTTCELYELVEAMVEKQDQSAVLELDLCNRDVSRTTFQKDCNKFTGTGTH 145
QY 1066 GNVGGGIAHKSILNVLAAWRIIEE-ILRTGSRVYKSNGLPDEE--EAMLEAKINQY 1123
DB 146 GNVGGGISAWSKTFPCLFGPWFRAIEKALIALLPQGVFYGDAPDDVFSAVAALAAASV 205
QY 1124 PHATVSAWTEFDTAHHNTSELFAALLERIGTPAAAVNLPRERCGKRTLRAGLGSV 1183
DB 206 -----FBNDFSEFDSTQNNFSLGLECAINEBCGMPQMLIRLY-----HLISAWTLQAP 254
QY 1184 VDGL-----LDSGAAWTPCRRNTIFSAAVMLTL--FRGVFAAFKGDLSL-LCGSHYLAPD 1235
DB 255 KESLKGFMWKHSGEPOTLLMTNTVMMNAVITHCYDFRDPQVAAFKGDLSIVLCSBYQSPG 314
QY 1236 AARLMGERYKTKHLKVEVQKIVPYIGLLVSAEQVTL--DPVYSA 1278
DB 315 AAVLIAGCGLK---LKVDFRPILGLVAGVVAAGLALPDVVRFA 355

RESULT 13

US-10-223-070-23
Sequence 23, Application US/10223070
Publication No. US20030109045A1
GENERAL INFORMATION:
APPLICANT: NELSON, RICHARD S.
APPLICANT: DING, XIN SHUN
TITLE OF INVENTION: RNA SILENCING SUPPRESSION
FILE REFERENCE: NLEL.006US
CURRENT APPLICATION NUMBER: US/10/223,070
CURRENT FILING DATE: 2002-08-16
PRIOR APPLICATION NUMBER: 60/313,185
PRIOR FILING DATE: 2002-08-17
NUMBER OF SEQ ID NOS: 24
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 23
LENGTH: 1187
TYPE: PRT
ORGANISM: Tobacco rat flea virus
US-10-223-070-23

Query Match 2.2%; Score 191.5; DB 15; Length 1187;
Best Local Similarity 17.3%; Pred. No. 7.9e-06;
Matches 198; Conservative 128; Mismatches 333; Indels 483; Gaps 42;
QY 65 AGHGVABEVLAHNFARGSTVLBIQPSLHSLAKLHGAPNAPVAD--YHGCTKYGTGDSR 122
DB 168 SGKQMSNELMWTDOFRKNI-----AANAVRCNNYTHGCTCRGSDGK 213
QY 123 HITLBSRSVATGRPEPRKADSLANGIASRTFCVDGSCAFYSRVGIANHSLYDTILE 182
DB 214 K-----GQVYALHSLYDFTLK 231
QY 183 EL-ANAFENHGLHNVRAFMHPEBLLYMDNVNNAELGYRFRVIEBPAAVKKCAFQGGDLR 241
DB 232 DLMATWYKTKTKVGHAAALFAPBSVL-----VDRCFL----- 263
QY 242 LHPFLDPIINESQERRIRBLAAGSYSR--AVIFSGDDMDGDAYLHDFTMLAVLYVN 299
DB 264 ---PSVD-----GYMKGNKITYFGPKDPSFTIHDWSEYKYLTK- 303
QY 300 YTPPE-GFSLHIEVORRHSSIELRITRAPGDRMLAVVPTS-----QG 343
DB 304 -FVSQGVNVFFEPQVQVGTMLPSIYR-----IAGVPRSSSGEYRRIYISRMEN 355
QY 344 LCRIPTNYVYDASGTEKTLTSGHKVNMILNFMQTPPEKELYDMTVMSPARALAI 403
DB 356 MVVPEIPDLVSTRELVEKCDLFVEKQFMDKCLDYIARLSQDLT--ISVVKSYLSNNVYL 414
QY 404 VVASEVTSSWNISPADIVRTVSLVYLHIERRAAVALVTAQDVGERTSPWESLGHV 463
DB 415 FINGAAVKKQSVDSRDLQLAQTL---LVKEQVAREVMEBLRAILTETVPTSLTDV 470
QY 464 LG-----SCCGLEMLKGTDPV-----TKRV 484
DB 471 LGLISRKMMKQPAKNIAGGVGMVGLIGFPPKVLVMAQDTPNGPELCYENSHKTKYI 530
QY 485 VDKTVVSLGDI-----ICD----- 499
DB 531 VFLSVVAIAGITLRRDIRDGLVKKLCMDPIDIKGAHVLDVENPCRYDINDPSSLYS 590
QY 500 -----VRLSPQV----- 507
DB 591 ASBSEETVLPDLSEVKAASDKLQKKEIADBFLSAKTSNYSGSSVTRSPSVVSSNSG 650
QY 508 -GFL-----PSRVPPARVPHREBLE----- 527
DB 651 LGLLEDSNVLTQAVGVSRKVADEIMEQFLSGLIDTEABIDVVPVAPASACERGETSG 710
QY 528 -----VIREACTYERP-----VSTPVEPQGFAD- 555
DB 711 TKVLGNLTLTPGFEENVLPVAVKPELVSKGTIVKVDYFQVWGBRLPKRPVPSGDDSDVAR 770
QY 556 --LMHATAASLPE-----YRATLQAGLNT-----DVKQLK-----ITLENA 589

```

Db      771 EFLVYLDAERVAQNDEIMSLYRDYSGRVIRTCGQNVPHGLGVWVDEMKNWCIRPVYTHEA 830
Qy      590 -----
Db      831 VYSNPKKMDMSGYLEVAVERGMLVNDFAVERNSDYIVCDQTYLCNNRLILDNLSL 890
Qy      597 TUSPVR-GELEMEYEGPGSGKTGTLIALEAAGKALYV--APTRELR-----AMD 644
Db      891 DLGPNCSFELVDVPGCGCKSTMIYNSANPCVDVVLSTGRATDILIERFSPSKGPCXIK 950
Qy      645 RRIKPPSASATQHVALATLRATAGCAPATV--IDECFMPLYVVAIVHALSPSSRIYL 703
Db      991 RRVKTVV-DSPLMH-----CVDGSLTGDVLHDEALMAGNVVCAQIAGAKRCIC 1000
Qy      704 VGDVHQIGF-----IDPQTSANMPLVR--DVKQCRRTFTNQKRCPADVAVATTPQOS 755
Db      1001 QGDQWQISFKPRVSOVDLRFSS---LVQKFDIVTEKR-----ETVRSBADVAAN--LNK 1049
Qy      756 LYPGCTTSGCVASISHVAPDVRNSO-----AQLLCTPOEKS-----RHGAEGM 801
Db      1050 YTTGVRTHNATANSMTVRKIVSKQVSLKPGAQYITPLQSEKELVNLALRKVAAYVS 1109
Qy      802 TVHNRGRTFASVILHYNGSTAEQKILAEKSHLVGTRPHTNHL-YIRDPGDIRQLNH 860
Db      1110 TVHSGQETFDVVLVVRTRKPT-DDSIARGREVLIVALSHTQSLVYETVWEDVSKETIRE 1166
Qy      861 SA 862
Db      1169 SA 1170

```

RESULT 14

```

US-09-801-368-108
; Sequence 108, Application US/09801368
; Patent No. US20020128250A1
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cai, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: No. US20020128250A1man, Thea
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
; FILE REFERENCE: 109272.147
; CURRENT APPLICATION NUMBER: US/09/801,368
; PRIOR FILING DATE: 2001-03-07
; PRIOR APPLICATION NUMBER: US 09/487,558
; PRIOR FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/160,587
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 440
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 108
; LENGTH: 1367
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-801-368-108

```

Query Match 2.1%; Score 186.5; DB 10; Length 1367;

Best Local Similarity 24.9%; Pred. No. 2.6e-05;

Matches 90; Conservative 56; Mismatches 177; Indels 39; Gaps 14;

Qy 1355 AHVQAPDAYSTTPYANVASCIDHVFEPROAAPAGFVATCAKPEPTP--SSITTAAGVSA 1412

Db 327 APVPTP--SSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVTSSTTESSAPV 383

```

Qy      1413 TTSVATGTAPESP-----WDAPANSFELLTPETSTSSSPSSSDSTSCGRS 1465
Db      384 TSSTTESSAPVPTPSSSTTESSAPVTSSTTE--SSAPVTSSTTESSAPVTSSTTES 441
Qy      1466 LSGDPTARTTEDLNSRKP-PEQDRQSRSECLDRGERTGSLTAPTAPEPSFSEPAR 1524
Db      442 SSAPVTSSTTESSAPVPTPSSSTTESSAPVTS--TTSSSAP-VPTPSSSTTESSS 497
Qy      1555 L-ANGPTVAAATSPGATPSCATDOVAARTPDPAFPLQSGARAVSKYRPPTTARKEV 1583
Db      498 APVTSSTTESSAPVPTPSSSTTE--SSAP--APTSSSTTESSAPVTSSTTESSAP 553
Qy      1584 TPLAMKVTDREVRNRPDEPTAIVVQALIGRPQTKLASDASKGSRKGC--QST 1641
Db      554 VPT-----PSSSTTESSSTPTSTSTES--SSAPVPTPSSSTTESSAPVPTPSSSTTES 607
Qy      1642 SPAPASADYQARDCTVAVCAALAEARSCIHBPPLASAPADL-----KRIRSTDS 1694
Db      608 SAPAPTPSSSTTESSAPVTSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESAP 667
Qy      1695 VP 1696
Db      668 VP 669

```

RESULT 15

```

US-09-815-242-5816
; Sequence 5816, Application US/09815242
; Patent No. US2002061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyekind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 1410
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5816
; LENGTH: 2478
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5816

```

Query Match 2.1%; Score 186.5; DB 9; Length 2478;

Best Local Similarity 18.8%; Pred. No. 7.3e-05;

Matches 260; Conservative 195; Mismatches 566; Indels 359; Gaps 55;

Qy 528 VLREAGCYNE-----RPVPTPV--EEPGFADL-----WHATASL 564

Db 1220 VNKATATNERTITALINNLGEGIQATPDATDEKQADABANTENGKANQATSAATTNNOV 1279

QY 565 PEYRATQAGLNT-----DYKOLITLENAKTIDGLTSPVROGLEMEGEP 611
 Db 1280 DEAKANAELAINAVTPKVVKKOAAKDEIDQLOATQTNVINDONATEEKEAIIQOLATA 1339
 QY 612 GSGKGTLLIALEAAG-----GKALVYAPTRLEBEAMDRRIKPPSASATOHVALAIR 664
 Db 1340 VTDAGNNITATDDNDGVDOAKAGAKNSIOSTQBPATVKNANQDVQATTTQNOALDNTT 1399
 QY 665 RATAEGAPPAIVIDECEMFPLVYVAIYHALSPSSRIVLVGDVHQI--GFIDPGTSAN 721
 Db 1400 GATTEBKNAK-----DVLKAKEKAYQDILNAQTNDVYTKQDAVADIQITAD 1450
 QY 722 MELVADVYKQCRRTTNTQKRCRADV-ATT-----FPGSLXP 758
 Db 1451 -TTIKVADKELATKANKEKALIAQTADATTEKEQANOQVDAQLTQGNONIEMASID 1509
 QY 759 GCTTSGCVASISHV--APDYRNSQATLCTQOEKSRHAGAGAMTVHBAQRTFASVIL 816
 Db 1510 VNTADNMAIQALDPIQASTDVKTNAABELLTEWOKITELINNETTEKNDIGPVA 1569
 QY 817 HY-----NGSTAEOKLIAEKSHLLVITR-HTNHLVIRDPGTIEROLNHSKAEVPT 868
 Db 1570 AYEGLNNINAAITTTGDTTAKDTAVQVQOLHAP--VKKPAG--KKEILDQA--ADKKT 1624
 QY 869 DI-----PAPLEITTVKPESEVQORNEVMAITIPQSATPHGAILHLKKNFGDQDCCVAL 923
 Db 1625 QIEQFPNMQOEINDAKQEVDTLNQAKTNV--DOSSTNEYVDNAKE----- 1670
 QY 924 AKTGEVFGGRAKINVELA---EPDATKPRHAFQEGVQWVVTASNKH---QALQTL 976
 Db 1671 -----GKAKINAVKTFSEYKCDALAKIEDAYNAKNEADNSASTSSIEAKOKL 1721
 QY 977 LSRVYKRSADLPLEHA--KEDVK-RMLNSLDRHMDVTEADARAIVFETQLKFTQSGT 1033
 Db 1722 AE--LKQTADQWVNGATSKODIEVQHNDLNDINDYTIPTGKKESATDLYAYADQKNN 1779
 QY 1034 VE-DLLEPD---PYRDIIDFLMKTOQKVSPPKINTGKVGCGIAMSXSINFLAAMIRI 1089
 Db 1780 ISADTNATODEKQOAKQVDQVQJALIESINGVDVDALDQKKAIDALQV----- 1834
 QY 1090 LBEILRTGSRVYSNGLPDEEBAMLLEAKINOVPHATFVSADWTEFDTAHNNTSELPA 1149
 Db 1835 -----DATVK-----PKAQAIEVKA-----DTKESIDQSLTAEKTEAL--- 1872
 QY 1150 ALLERIGTPAAAVNLFRERCGRKTLRAKGLGVEVDGLDSCAAWTPCRRITJPSAAVMT 1209
 Db 1873 AMIKQI-----TDQAK-----QGITDA-----T 1890
 QY 1210 LFRGVYFAAFKGDSDILCGSHYLFPAASRLHMERKYT-KHLKVEVQKIVPIGLVSAE 1268
 Db 1891 TTAEVKAKAQAQIEA-----FDNIQIDSTEKQKALIELETALDQIEAGVNAVADAT 1941
 QY 1269 QVULDEVRSAKIFGRCTSELLYSKYVAVRDITKWSMDARX-HSLLCHMSACYNYAP 1327
 Db 1942 TEEKEAFNAL-----EDILSKATEDISDQTNMAELATKNSALEOKAKORIN--P 1990
 QY 1328 ESAAYIIDA-----VRFGRGDFP-EQLRVVRAHVQAPD 1361
 Db 1991 EYKXNALBAIREVVNQIEIKVADASAKELAFDTLGRYFDRPADKLDKQTNAEVAE 2050
 QY 1362 AYSSTYPA-----NVRASCLDHVEFRQAAAAPAGFVATCAKPEP--SLITA--KAG 1409
 Db 2051 LQNVITPAIEAIVPQNDPANDTNGIINNDAFANSANATPENTQOPNVSETTANGKAD 2110
 QY 1410 VBAAT--SHVATG---TAPESPMDAPAPANSFSELLTPEPTSSSPSSSSSDSTSC 1462
 Db 2111 ASPTTNNSDAATGETTATSATDANDKQANNSSV-----DASTNSPTMDNDVTSKE 2165
 QY 1463 GSSLGSGDJARTTEDINSRPPSQDRQSRSSSECLDRSGERTGSSLTAPAPSPSPSPSER 1522
 Db 2166 VESTINGTIDKPVTEFDNATPAESTTNNSTTTATNENAPYTGSTATAPITAS-----TEA 2220

QY 1523 ARLATGPTVA-----AATSPATPSCATQVA-----ARTTPDPAFLGSGSARA 1567
 Db 2221 ASSADSKDNASVNSKQNAEYVNSHESQSTNDKVAQPSKAKAKKD-----GSDSTNQ 2275
 QY 1568 -----VSKYRPPTTARMB-----VTPHAMKGVTCGR 1596
 Db 2276 SMVSTTETLPBADITBPNVPENTSKQEBSTTNGTDAGQLKSEFTNVASNEADKSPSKAD 2335
 QY 1597 PEYREDPEYAAVQA---LISGRYOK-----TKLSQDASKYSTRKGCQOST 1641
 Db 2336 TEVSXKPFYSASSEAKERTSTNVQKODTATADTNDTQKSVGSAANNKATQNDGANASP 2395
 QY 1642 SPAPSAVQARDCCQTVRCRAAAMASCIHEPLASASADLKRIBSTSDSVPOVKIS 1701
 Db 2396 ATVSNGSNANQDMNV-----TNTDHOAKTKSHQGGKVNKAQOAKTLPDTGNS 2446

Search completed: January 15, 2004, 17:23:16
 Job time : 215.393 secs

XX Modified small RNA viruses and virus-like particles - have altered
 PT or substituted Ig-like domains to modify host cell tropism, useful
 PT as insecticides and in medicinal applications
 XX
 XX
 PS Disclosure: Figure 1, 41pp; English.

XX The sequence is that of coat protein p71, it has an Ig-like domain
 CC and can be used in the production of virus-like particles (VLP). The
 CC VLPs can be used in vaccines where the Ig-like domain has been altered
 CC so that the VLP presents a surface located antigen which is used to
 CC elicit an immune response in a host organism. They can also be used
 CC for controlling the proliferation of a pest insect and potentially as
 CC medicinal delivery agents for cancer treatment and gene therapy.

XX Sequence 647 AA;

Query Match 100.0%; Score 3374; DB 19; Length 647;
 Best Local Similarity 100.0%; Pred. No. 4,2e-273; Mismatches 0; Indels 0; Gaps 0;
 Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDAVASQRPNNRGTNRVNSANTVTYVNGRRNRRTGRGROVSPDNFTAAQDLAOSL 60
 DB 1 MGDAVASQRPNNRGTNRVNSANTVTYVNGRRNRRTGRGROVSPDNFTAAQDLAOSL 60
 QY 61 DANTVTPPANTISSMEPEFNMAKGIIDSDSIGWYFKYLDPAGATESARAAGEYSKIPDG 120
 DB 61 DANTVTPPANTISSMEPEFNMAKGIIDSDSIGWYFKYLDPAGATESARAAGEYSKIPDG 120
 QY 121 LKFSVDVAEIRIYNEECPVVTVDSVPLDGRQMSLSISFPFRFAVAVANVENKMSL 180
 DB 121 LKFSVDVAEIRIYNEECPVVTVDSVPLDGRQMSLSISFPFRFAVAVANVENKMSL 180
 QY 121 LKFSVDVAEIRIYNEECPVVTVDSVPLDGRQMSLSISFPFRFAVAVANVENKMSL 180
 DB 121 LKFSVDVAEIRIYNEECPVVTVDSVPLDGRQMSLSISFPFRFAVAVANVENKMSL 180
 QY 181 DVNDLIMLNINLADMRVYVDSBQWINFNTDTTYVIRVLRPTVDVDPTEGLVRYTSD 240
 DB 181 DVNDLIMLNINLADMRVYVDSBQWINFNTDTTYVIRVLRPTVDVDPTEGLVRYTSD 240
 QY 241 YRLTYKAITCEANMPTLVDOGFWMIGQYALPTSLPQYDVSBAYALHTLTFARPSSAAL 300
 DB 241 YRLTYKAITCEANMPTLVDOGFWMIGQYALPTSLPQYDVSBAYALHTLTFARPSSAAL 300
 QY 301 AFVWAGLPQGGTAPAGTAMWQASGGYLTWRHNGITTPAGSVSYLPEGFALERYDPND 360
 DB 301 AFVWAGLPQGGTAPAGTAMWQASGGYLTWRHNGITTPAGSVSYLPEGFALERYDPND 360
 QY 361 GSWTDFASAGPTVTRQYAVDEVVVTNNPAGGSAPTFTVRYPSPNAATNTVFRNTLLET 420
 DB 361 GSWTDFASAGPTVTRQYAVDEVVVTNNPAGGSAPTFTVRYPSPNAATNTVFRNTLLET 420
 QY 421 RPSRRLELPMPADFGQTVANNPKIEQSLKETLGCYLVHSGKMPVQOLTTPASSFGAV 480
 DB 421 RPSRRLELPMPADFGQTVANNPKIEQSLKETLGCYLVHSGKMPVQOLTTPASSFGAV 480
 QY 481 SNNNGYERTRLPYTGIRDSFDQMSYAVHSPSLSHSGSIYTKTQMGVNTVMT 540
 DB 481 SNNNGYERTRLPYTGIRDSFDQMSYAVHSPSLSHSGSIYTKTQMGVNTVMT 540
 QY 541 FQGFPAHAGLKNKEELICLADDLATRLTGYPATNDPFAAASFAANMLSSVLSKESATSI 600
 DB 541 FQGFPAHAGLKNKEELICLADDLATRLTGYPATNDPFAAASFAANMLSSVLSKESATSI 600
 QY 601 IKSVEVAVGAAGSLAKLPGLLMSVPCKIARVARRARRRABAN 647
 DB 601 IKSVEVAVGAAGSLAKLPGLLMSVPCKIARVARRARRRABAN 647

RESULT 2

ID AAR49662 standard; Protein; 647 AA.

XX AAR49662;

DT 25-MAR-2003 (updated)

DT 12-SEP-1994 (first entry)

DB Sequence of Heliothis armigera RNA 2 P17.

KW HaSV; RNA 1, small RNA virus, P17.

XX Heliothis armigera stunt virus.

XX WO9404660-A1.

XX 03-MAR-1994.

XX 13-AUG-1993; 93WO-AU00411.

XX 14-AUG-1992; 92AU-0004081.

XX 08-JUL-1993; 93US-0089372.

XX (CSTR) COMMONWEALTH SCI & IND RES ORG.

XX (PACT-) PACIFIC SEEDS PTY LTD.

XX Christian PD, Gordon KHJ, Hanzlik TN;

XX WPI, 1994-083180/10.

XX N-PEDB; AAQ58523.

XX Small RNA virus capable of infecting insect species, e.g.

XX Heliothis - and transgenic plants contg. viral nucleic acid, for

XX protection against insect pests

XX Disclosure: Figure 2, 183pp; English.

XX The inventors claim a virus comprising a genome hybridisable with

XX the nucleotide sequence of RNA 1 or RNA 2; pref. the sequences are

XX those given in Figs 1 and 2 of the specification. As isolated

XX protein or polypeptide prepn. of the proteins or polypeptides

XX H. armigera larvae were raised and viral RNA was extracted. The virus

XX RNA were reverse transcribed into cDNA. Clone hr236 contains about

XX 88% or RNA 2. A major translation product of apparent mol. wt. 24,000

XX is obt. This protein is derived from a mol. wt. 17,000 reading

XX frame overlapping the abla of the capsid protein gene. The Mr 24,000

XX protein (referred to as P17) may have a function in modifying or

XX manipulating the growth characteristics or cell cycle of

XX CC (HasV)-infected cells.

XX (updated on 25-MAR-2003 to correct PN field.)

XX Sequence 647 AA;

Query Match 99.4%; Score 3353; DB 15; Length 647;

Best Local Similarity 99.5%; Pred. No. 2.4e-271;

Matches 644; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MGDAVASQRPNNRGTNRVNSANTVTYVNGRRNRRTGRGROVSPDNFTAAQDLAOSL 60
 DB 1 MGDAVASQRPNNRGTNRVNSANTVTYVNGRRNRRTGRGROVSPDNFTAAQDLAOSL 60
 QY 61 DANTVTPPANTISSMEPEFNMAKGIIDSDSIGWYFKYLDPAGATESARAAGEYSKIPDG 120
 DB 61 DANTVTPPANTISSMEPEFNMAKGIIDSDSIGWYFKYLDPAGATESARAAGEYSKIPDG 120
 QY 121 LKFSVDVAEIRIYNEECPVVTVDSVPLDGRQMSLSISFPFRFAVAVANVENKMSL 180
 DB 121 LKFSVDVAEIRIYNEECPVVTVDSVPLDGRQMSLSISFPFRFAVAVANVENKMSL 180
 QY 121 LKFSVDVAEIRIYNEECPVVTVDSVPLDGRQMSLSISFPFRFAVAVANVENKMSL 180
 DB 121 LKFSVDVAEIRIYNEECPVVTVDSVPLDGRQMSLSISFPFRFAVAVANVENKMSL 180
 QY 181 DVNDLIMLNINLADMRVYVDSBQWINFNTDTTYVIRVLRPTVDVDPTEGLVRYTSD 240
 DB 181 DVNDLIMLNINLADMRVYVDSBQWINFNTDTTYVIRVLRPTVDVDPTEGLVRYTSD 240
 QY 241 YRLTYKAITCEANMPTLVDOGFWMIGQYALPTSLPQYDVSBAYALHTLTFARPSSAAL 300
 DB 241 YRLTYKAITCEANMPTLVDOGFWMIGQYALPTSLPQYDVSBAYALHTLTFARPSSAAL 300
 QY 301 AFVWAGLPQGGTAPAGTAMWQASGGYLTWRHNGITTPAGSVSYLPEGFALERYDPND 360

```

Db 301 AFVAGLPGGTAGTAPGAWQASGGYLTRHNGTTPASVSIVLREGALERYDND 360
Qy 361 GSWTDFASAGTTFTRQVAVDEVVVTNNPAGGSGAPFTTVPPSNAYTNTVFRNTLIET 420
Db 361 GSWTDFASAGTTFTRQVAVDEVVVTNNPAGGSGAPFTTVPPSNAYTNTVFRNTLIET 420
Qy 421 RPSRRLELMPMPADFGQTVANNPKIEBSLKEITGCVLVSKRMNVPOLTPASSPGAV 480
Db 421 RPSRRLELMPMPADFGQTVANNPKIEBSLKEITGCVLVSKRMNVPOLTPASSPGAV 480
Qy 481 SFNNPGYERTDLDPDYTGIRDSFDONMSTAVAHFRLSHSCSIYTKTYQMEGTVNTP 540
Db 481 SFNNPGYERTDLDPDYTGIRDSFDONMSTAVAHFRLSHSCSIYTKTYQMEGTVNTP 540
Qy 541 FGOFAHAGLKNKEITLCLADLALTRNGVYPATDNFAAASAPAAAMLSSYLKSEATSI 600
Db 541 FGOFAHAGLKNKEITLCLADLALTRNGVYPATDNFAAASAPAAAMLSSYLKSEATSI 600
Qy 601 IKSVGETAVGAAGSLAKLPGILMSVPGKIAARVAPARRRRAAAN 647
Db 601 IKSVGETAVGAAGSLAKLPGILMSVPGKIAARVAPARRRRAAAN 647

RESULT 3
ID AAW26785 standard; Protein: 634 AA.
AC AAW26785;
XX 22-JUN-1998 (first entry)
DE Nudaurelia beta-like virus capsid protein precursor.
XX NBV; RNA virus; transgenic plant; insect resistance;
XX disease resistance; Nudaurelia cytherea capensis;
XX pine tree emperor moth; capsid protein; vector;
XX virus-like particle.
OS Nudaurelia beta-like virus.
XX AAW26785-A.
XX AU9724669-A.
XX 04-DEC-1997.
XX 02-JUN-1997; 97AU-0024669.
XX 31-MAY-1996; 96AU-0000233.
XX (CSIR ) COMMONWEALTH SCI & IND RES ORG.
XX (UTRH-) UNITV RHODES.
XX Gordon KH, Hanzlik TN, Hendry DA;
XX WPI: 1998-052736/06.
XX DR N-PSDB: AAV04471.
XX PT Nudaurelia beta virus nucleic acid - useful for producing
XX recombinant virus, insect-resistant transgenic plants, etc
XX Example 1; Fig 1; 33pp; English.
XX This polypeptide comprises a 70 kDa capsid protein precursor whose
XX amino acid sequence was deduced from an open reading frame identified
XX in the RNA genome (see AAV04471) of Nudaurelia beta-like virus (NBV).
XX The capsid protein precursor is cleaved at a N/G site into 60.5 and
XX 8 kDa capsid proteins. A claimed infectious recombinant insect virus
XX vector comprises an expressible nucleic acid molecule comprising a
XX nucleotide sequence corresponding to all or an infectious and/or
XX transgenic plant resistant to insect attack that produces NBV such
XX that insects feeding on the plant are deleteriously affected; and a
XX virus-like particle (VLP) prepared from expression of a nucleic acid

```

```

CC molecule comprising a sequence encoding the capsid protein of NBV.
CC The invention provides methods for the control of insect pests (the
CC pine-tree emperor moth Nudaurelia cytherea capensis is mentioned).
XX Sequence 634 AA;
SQ
Query Match 8.6%; Score 289; DB 19; Length 634;
Best Local Similarity 25.7%; Pred. No. 3.9e-15;
Matches 177; Conservative 72; Mismatches 266; Indels 174; Gaps 39;
Qy 35 QRRRTGRVSPDPFTAAQ--DLAQSLDANTVFPANISGM-----PS----- 76
Db 31 QRAKTEKQPEPPATAAQTITTSIQBAGSKTSPRSRTDYQAPAMNPPRRHPPQ 90
Qy 77 --FNNMAKXKIDLD--SDSIGWYFKYLDAPAGATASARAVGEYSKIDGLVKSVDAR 131
Db 91 SDTRGKAASDDGSHSDIKAMHNDYLDPDGBEKTSLDDQ--KIPGALPGTCGQFR 147
Qy 132 EIVNBECPVTVDPVLPDGRQMSLISFPMFRFAVAVANVENKMSLDVYN-DL---- 186
Db 148 GTVGARYFGALNSTTLPDGGTWPLVMHLPEFRHPLFTTTSMT--VEVTNADLDAFA 205
Qy 187 IEWLNLDNRXYVVDSEQWTFNDITYVYAIRVLRPT--YDVPDPTL-GLVTVSDYR 242
Db 206 NDW--NNRTDWTET--YPSMAQVGN--VFYVVV--PTBALTDVPPPTQLGVSGLBSYR 257
Qy 243 LTYKAITCEANMPTLVDOGFITGGQYALTPSLPQYDVSBAVALHTLT-----FA 295
Db 258 LTSQGVAYFNAFPLVNGVAVIAQF--QPKKHQKKNPDIVAQTGTGTLQLGSGGPN 315
Qy 296 SAALAFVMAGLPGGFTA-PAGTPAWQASGGYLTRHNGTTPAG--SVSYVLPGEF 351
Db 316 YTLTMT-IGDOVERGAAHPLPTVSMGPBESGQLVQTNALTFDVGNTITTTLP 374
Qy 352 ALERYDPRDGSWTFASAG-DTVTFQVAVDEVVVTNNPAGGSGAPFTTVAPSNAYTN 410
Db 375 V-----TGMMQGTASNGTDTV--VD-----AGA-----TVAV----- 400
Qy 411 TVPRNTL--LETSSRLSL--PMPADFGQTVANNPKIEBSLKEITGCVLVSKM 464
Db 401 ---RSBFRGLTBSAGQLNDSTNDMNPNDAG---NAKTIQOLTKR--GHYMPKASI 451
Qy 465 RNPVFOITPASSFGAVFPN-----PGYER--TRD-LDPYTGIRDSFDONM 507
Db 452 R---VRNDNATSYVPVSDTBDCGRHLRAIGLQGYHRQQLRRLPSMTG-----M 500
Qy 508 STAVAHFRSLSHSCSIYTKTYQMEGV-----TNVTPFGQFAHAGLKNKEITCLAD 561
Db 501 STSTVPY-----LQGVPTLRSDTGGGEPMPGFASATPPRDVALLVART 544
Qy 562 LATRLTGVPATDNFAAASAPAAAMLSSYLKSEATSIKSVGETAVGAAGSLAKLPG 621
Db 545 WTDLHPAYPERYNGFALFAMVAKTIQAQIPR-----YVRS---AAGVANAVTDCIES 594
Qy 622 LMSVPGKIAARVAPARRR-----RAAR 645
Db 595 ATBSVASNSTSERKRRARRRGVGIARGAR 623

RESULT 4
ID AAW34537 standard; Protein: 634 AA.
AC AAW34537;
XX 08-JUN-1998 (first entry)
DE Nudaurelia beta virus coat protein p70.
XX Vaccine; coat protein; p70; insecticide; Ig-like domain.
XX Nudaurelia beta virus.

```

Key Location/Qualifiers
 285..433
 /note="Immunoglobulin-like domain"

MO9746666-A1.
 11-DEC-1997.
 02-JUN-1997; 97WO-AU00349.
 31-MAY-1996; 96AU-0000234.
 (CSIR) COMMONWEALTH SCI & IND RES ORG.
 Gordon KH, Hanzlik TN;
 WPI, 1998-042175/04.
 N-PSDB; AAT99118.

Modified small RNA viruses and virus-like particles - have altered
 or substituted Ig-like domains to modify host cell tropism, useful
 as insecticides and in medicinal applications

Disclosure; Figure 2; 41pp; English.

The sequence is that of coat protein p70, it includes an Ig-like domain
 which can be used in the production of virus-like particles (VLPs). The
 VLPs can be used in vaccines where the Ig-like domain has been altered
 so that the VLP presents a surface located antigen which is used to
 elicit an immune response in a host organism. They can also be used
 for controlling the proliferation of a pest insect and potentially as
 medicinal delivery agents for cancer treatment and gene therapy.

Sequence 634 AA;

Query Match 8.6%; Score 289; DB 19; Length 634;
 Best Local Similarity 25.7%; Pred. No. 3.9e-15;
 Matches 177; Conservative 72; Mismatches 266; Indels 174; Gaps 39;

35 QRRRTGRVSPPDNFTAAQ--DLAQLDANTVTPANISSM-----PE-----76
 31 QGADKTRKQEPFRAQTTRTTSTGAGSKTSRSTVDYQAPMPBPREHGPQR 90
 77 --FRWMAKIDLD--SPSICWYRKYLDPAQATSAKAVGSKIPDGLVPSDAIR 131
 91 SDTRGAKASDGEHSGDIKAMIDHYLDPDGEYKTSLDG--KIPDAIQSTCGQPR 147
 132 ELYNECPVYVDVSVPLDGRKMSLSPFPMFRTAVVAANVENKMSLDVYN-DL----186
 148 GTVGARYPGINSTTLPDGGTWPVLVHMLPFRHPLFTTTTSNTE--VEVTNADLDAPA 205
 187 IEMLNLDMDRYVVDSEQWINFNTDITYVRIKVLPT--YVDPDPTG-GLKRTVSDYR 242
 206 NDW-NNRDWTREAT-YPSWAQVGN--VFYMWV---PTEALDVPPTQLGVSGLLESYR 257
 243 LTYKAITGANNPTLYDQFMIGGQYALFTPSLPQYDVEAVALHTLT-----PARS 295
 258 LTHSGTAYFNAPITLVNCGVAVIAQF--OPDKHOKENDIVAGTQGTGLQLGSGSGN 315
 296 SAALAFVWAGLPOGGA-PAGTPAWEQASGGCYLWRNNGTTPFPAQ---SVSYVPEGR 351
 316 YTLTMT-IGDQVBFQGAALPLPTVSMGPMPSGQLVFOYANLTFDVGNTTTTTLPPPS 374
 352 ALERYDPNDGSWTDFASAG-DTVTFRQVAVDEVVYTNPNAGGSAFTFVRVPPSNAYTN 410
 375 V-----TGMQWFTASNGTDITV-----VD-----AGA-----TVRV-----400
 411 TYFRNTL--LETRPSSRLLEL-----PMFADRGQYVANNPKIEGSLKTKGLYVHSGM 464
 401 ---RSFGRGLGESASHQLNQDSTNDMPNDAG---NAKTIQFOLTKK--GHYMEASAI 451
 465 RAPVFOLTPASSFGAIVFNN-----PGEYR--TRD-LPDVTGIRDSFDQNM 507

DB 452 R---VRNDNATSGVPVDEDTEDNGRLHRAIGALQGYHRQQLRDRLPBMTG-----M 500
 QY 508 STAVAHFRSLHSGSIVTKTYQMBGV-----TNVTPFQGFAPHAGLLKNEBILCLADD 561
 DB 501 STSTVVPY-----LQGVPTLRSDPGGGFPMQPFASATPPKODVALTVART 544
 QY 562 LATRLTGVPATDNPAAVSAFANMLSSVLSKRSATSSIIISVETAVNGAQSGLATKRG 621
 DB 545 WTDLHPFAYPERRYNGFGLFPMVAKTTIAQIR-----YNS---AAGVANAATVDCIES 594
 QY 622 LMSVPGKIAARVARRAR-----RAAR 645
 DB 595 ATBSVANSSTSRRRQRARRRVGIARGAR 623

RESULT 5
 AAY00219
 ID AAY00219 standard; Protein, 1231 AA.

AAV00219;
 20-APR-1999 (first entry)

Enterococcus faecalis antigenic polypeptide fragment EF108.
 Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 detection; attenuation; antigenic.
 Enterococcus faecalis.
 MO9850554-A2.
 12-NOV-1998.
 04-MAY-1998; 98WO-US08959.
 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 (HUMA-) HUMAN GENOME SCI INC.
 Bailey C, Choi GH, Hromocky J A, Kunsch CA;
 WPI, 1999-070095/06.
 N-PSDB; AAX20209.

New isolated Enterococcus faecalis polynucleotides - used to develop
 products for the detection of Enterococcus and for use in vaccines
 for prevention or attenuation of Enterococcus infection

Claim 9, Page 211, 301pp; English.

The present sequence represents an antigenic polypeptide fragment
 isolated from Enterococcus faecalis. The present invention describes
 genes, proteins and antigenic polypeptides isolated from E. faecalis.
 The proteins can be used in vaccines for preventing or attenuating an
 infection caused by a member of the Enterococcus genus in an animal.
 CC They can also be used for detecting Enterococcus antibodies in a sample.
 CC The nucleotide sequences can be used for detecting Enterococcus nucleic
 acids. Products from the present invention can also be used for
 screening compounds to identify agonists and antagonists of E. faecalis
 protein activity.

Sequence 1231 AA;

Query Match 4.1%; Score 139.5; DB 20; Length 1231;
 Best Local Similarity 19.0%; Pred. No. 0.034; Indels 257; Gaps 31;
 Matches 141; Conservative 94; Mismatches 252;

19 NTVRSANTVTVNGRNRGRRTGRQVSPPDNFTAAQDLAQSIDANTV---TFPANISS-73
 542 NRKVTENFVDNIGAK-----ITPPTGFTQGGKVTYTS-DATYFQAGTLPDYTTTG 591


```

QY 74 --MPEFNNMAKGIKIDLSDSIGWFKYLDPAAGATESARAAGEYSKIDGLYKESVDNAIR 131
DB 592 GKTYYKFGWKYSILMTLT-----TKAPSYOYTYDNDNDLN 629
QY 132 EYNEECPPVTVDSVPLD-----GROMSLSIFAPMERTAYVAANYEN 175
DB 630 VYEEETVTVYVVDNMFVNEKGAFTPALTFSGKTYAOSTSAV--LRTDLYVYTSKNN 687
QY 176 -----KEMSLDVVNDLIEWLN-----191
DB 688 GNGQYTVSINNNSWPLSQELLKKYNNNGPISATNRLQFNVDKLAIDQLKYVDSIQDPTA 747
QY 192 ---NLAMRYV-----DSEOMINTNDTTYV---RIRVLAR 223
DB 748 QSSNLKSKYRYYYTNNSSLVDPDPNVAPEAVDLSSSLMLNLDSDGYFSSNANNRLFYTHL 807
QY 224 TYD-VDPDTBGLVYTVSD-----YRLTYRAITCEAMPTLVNQGFWIGQVALTP 272
DB 808 GYSGTPEGVNTLVNFWFLNNAKPADSKLYKVKTRKQVT-----ENFVD---VNGAKITAP 858
QY 273 TSLPQYDVSEAYALHTLTFARPSSAAL-----AFVWAGLPQGGTAPAGTPAMEQA 323
DB 859 TGFQY---GNGOVNNSMTF--KYTAAKALPATYTTGKAYTFQGWYKGTKSTL-----908
QY 324 SSGGYLTWRHNGTTFPAGSVSYVLPBGFALERYDPNDGSKTDPASAGDTVTFROVAVDEV 383
DB 909 -----NKITTFENAT-----PDGNDMTAMYKEIPTASVTLTRPEXV 947
QY 384 VVTNNPAGGSAFTFYRVPPSNAYNTVTFRNTLLETTPS--SRRLBEP--NPPADFGQTV 440
DB 948 IDNTNVIWTTTNTTNSKAP-----LQNLTLKGKPNMSAGLITPTFMEVTPBGETT 998
QY 441 ANNPKIBOSILKETLGCYLVHSSKNRNPVFOULTPASFGAVSF-----NNPGYERTRLDLP 495
DB 999 KSLP-VNSTLMTBEGV-----PLPNAVPFGKKVSVAFTRATGKPTVTKAEVVV 1046
QY 496 YTGIRDSFDQNMSTAFAHFRSLSHSCSYVTYTGQMBGVNTVNTPPQGFAPAGLKKKEBI 555
DB 1047 FGGIKDSITVDNF-----VRIKPDQEVPTTGGFISVPTFD--FGQVGAAGTKQOHS 1098
QY 556 LCLAD-----DLATRLTGYPATDNF-----AAVASAPAN 586
DB 1099 KQADVYNGNTRNRYLAIKKTQPMWSLTAQLSQKSAIDSLPTATRLLLGAAPVSSFTNY 1158
QY 587 MLSVLKSE--ATSSIIKSVGETA 608
DB 1159 NQPTLKNYGTTSALISLTANNTA 1182

```

```

XX
PI Choi GH, Bailey C, Hromockyj A, Kuneach CA,
XX WPI; 2002-425450/45.
DR N-PSDB; ABN98194.
XX
PT New genes and polypeptides from Enterococcus faecalis, useful as
PT vaccines for preventing, treating or attenuating an infection caused by
PT a member of the Enterococcus genus in an animal, particularly E.
PT faecalis.
XX
PS Claim 9; Page 196; 255pp; English.
XX
CC The present invention provides the protein and coding sequences of a
CC number of polypeptides from Enterococcus faecalis. The proteins can be
CC used as vaccines for preventing or attenuating an infection caused by a
CC member of the Enterococcus genus in an animal, particularly E. faecalis.
CC The polynucleotide is also useful for preventing or treating E. faecalis
CC infection. The present sequence is a protein of the invention.
XX
SQ Sequence 1231 AA;
Query Match 4.1%; Score 139.5; DB 23; Length 1231;
Best Local Similarity 19.0%; Pred. No. 0.034;
Matches 141; Conservative 94; Mismatches 252; Indels 257; Gaps 31;
QY 19 NVKVSANTVYNGRRNRRTGRQVSPDNFTAAQDLAQLDANTV---TPPANISS- 73
DB 542 NRKVTENFVDNTGAK-----ITPPTGTOQKKTIVTS-DAYTFKQAGTLPDITYTG 591
QY 74 --MPEFNNMAKGIKIDLSDSIGWFKYLDPAAGATESARAAGEYSKIDGLYKESVDNAIR 131
DB 592 GKTYYKFGWKYSILMTLT-----TKAPSYOYTYDNDNDLN 629
QY 132 EYNEECPPVTVDSVPLD-----GROMSLSIFAPMERTAYVAANYEN 175
DB 630 VYEEETVTVYVVDNMFVNEKGAFTPALTFSGKTYAOSTSAV--LRTDLYVYTSKNN 687
QY 176 -----KEMSLDVVNDLIEWLN-----191
DB 688 GNGQYTVSINNNSWPLSQELLKKYNNNGPISATNRLQFNVDKLAIDQLKYVDSIQDPTA 747
QY 192 ---NLAMRYV-----DSEOMINTNDTTYV---RIRVLAR 223
DB 748 QSSNLKSKYRYYYTNNSSLVDPDPNVAPEAVDLSSSLMLNLDSDGYFSSNANNRLFYTHL 807
QY 224 TYD-VDPDTBGLVYTVSD-----YRLTYRAITCEAMPTLVNQGFWIGQVALTP 272
DB 808 GYSGTPEGVNTLVNFWFLNNAKPADSKLYKVKTRKQVT-----ENFVD---VNGAKITAP 858
QY 273 TSLPQYDVSEAYALHTLTFARPSSAAL-----AFVWAGLPQGGTAPAGTPAMEQA 323
DB 859 TGFQY---GNGOVNNSMTF--KYTAAKALPATYTTGKAYTFQGWYKGTKSTL-----908
QY 324 SSGGYLTWRHNGTTFPAGSVSYVLPBGFALERYDPNDGSKTDPASAGDTVTFROVAVDEV 383
DB 909 -----NKITTFENAT-----PDGNDMTAMYKEIPTASVTLTRPEXV 947
QY 384 VVTNNPAGGSAFTFYRVPPSNAYNTVTFRNTLLETTPS--SRRLBEP--NPPADFGQTV 440
DB 948 IDNTNVIWTTTNTTNSKAP-----LQNLTLKGKPNMSAGLITPTFMEVTPBGETT 998
QY 441 ANNPKIBOSILKETLGCYLVHSSKNRNPVFOULTPASFGAVSF-----NNPGYERTRLDLP 495
DB 999 KSLP-VNSTLMTBEGV-----PLPNAVPFGKKVSVAFTRATGKPTVTKAEVVV 1046
QY 496 YTGIRDSFDQNMSTAFAHFRSLSHSCSYVTYTGQMBGVNTVNTPPQGFAPAGLKKKEBI 555
DB 1047 FGGIKDSITVDNF-----VRIKPDQEVPTTGGFISVPTFD--FGQVGAAGTKQOHS 1098
QY 556 LCLAD-----DLATRLTGYPATDNF-----AAVASAPAN 586
DB 1099 KQADVYNGNTRNRYLAIKKTQPMWSLTAQLSQKSAIDSLPTATRLLLGAAPVSSFTNY 1158

```

QY 587 MLSSVLKSE--ATSSIIKSVGETA 608
 DB 1159 NQPTLKNVTGTTSAISLTANNTA 1182

RESULT 7
 ABU13717
 ID ABU13717 standard; Protein; 1231 AA.
 AC ABU13717;
 XX
 DT 26-FEB-2003 (first entry)
 DE Enterococcus faecalis EF040 polypeptide #210.
 XX
 KM EF040; immunostimulant; antibacterial; gene mapping.
 XX
 OS Enterococcus faecalis.
 XX
 PN US6448043-B1.
 XX
 PD 10-SEP-2002.
 XX
 PF 04-MAY-1998; 98US-0071035.
 XX
 PR 06-MAY-1997; 97US-044031P.
 PR 16-MAY-1997; 97US-046655P.
 PR 14-NOV-1997; 97US-066009P.
 PR 14-NOV-1997; 97US-066009P.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;
 DR N-PSDB; ABX61764.
 XX
 PT New EF040 polypeptides and polynucleotides from Enterococcus faecalis,
 PT useful for generating an immune response against E. faecalis and other
 PT Enterococcus species, and as vaccines against other bacterial genera -
 XX

PS Example 1; Column 217-220; 146pp; English.
 XX
 CC The invention relates to polynucleotide fragments of a gene from
 CC Enterococcus faecalis, EF040, and the polypeptides encoded by them. The
 CC polypeptides are useful in detecting E. faecalis, as epitope tags, as
 CC molecular weight markers on SDS-PAGE gels or for molecular sieve gel
 CC filtration columns, in generating antibodies that specifically bind to
 CC the E. faecalis polypeptides, in generating an immune response against E.
 CC faecalis and other Enterococcus species and as vaccines against other
 CC bacterial genera. The polynucleotides are useful as probes for gene
 CC mapping and for identifying E. faecalis in biological samples. Sequences
 CC ABU13508-ABU13755 represent EF040 polypeptides of the invention. Sequences
 CC Note: The sequence data for this patent can also be obtained from USPTO
 CC at seqdata.uspto.gov/sequence.html.
 XX
 SQ Sequence 1231 AA;

Query Match 4.1%; Score 139.5; DB 24; Length 1231;
 Best Local Similarity 19.0%; Pred. No. 0.034; Mismatches 252; Indels 257; Gaps 31;
 Matches 141; Conservative 94;

QY 19 NVKVSANTVTNGRRNRQRRRTGROVSPDNFTAAQDLAQSLDANTV---TFPANISS- 73
 DB 542 NKKVTENFVDTGAK-----ITPFGTGGKKTIVTS-DAYTFKAGTLPDVYTTNG 591
 QY 74 --MPEFRNAGKIDDSIGMYFYKLPDAGATESARAVGEYSKLPDGLVYFSVDAER 131
 DB 592 GKYRKKGWYKGSINTLTLT-----TKAPSYQVYTDNDL 629
 QY 132 EIVNECPVVTDSVPLD-----GRWSLSIFSFMPRTAVVAVANVEN 175

DB 630 VYIERETVTVPSVDNPNVNEKGAFTPALTPSGKYTAOSTSAV--LRTLDLYDYSKRN 687
 QY 176 -----KMSLDVNDLIEMLN----- 191
 DB 688 GNGQYTVSINNQSMPLSQGLKKNNGQPISATNRLGPNVDKLAIDQLKYVDSIQLDPA 747
 QY 192 ---NLADRRYV-----DSBQWINTNDYTYV---RIRVLAR 223
 DB 748 QSSNLKSYRYTYNNSGLVPDNPVAPAEVDLSSESLNLTNPDSDTYPSNANRLFYTHL 807
 QY 224 TYD-VPDPTBGLVTVSD-----YRLTYKAITCEANMPLVDQGFVIGQYALTP 272
 DB 808 GYSGTGPVNYLLVWELFPAKPADSKLYKTRKQVY-----ENFVD---VNGAKIARP 858
 QY 273 TSLPGYDVSBAVYLLTLPARPSSAAL-----AFVWAGLPQGGTAAPGTPAWBOA 323
 DB 859 TGFTQ---GNGVPMNSNPF-KYTAKALPATYTGKRYTFQGWYKGTKDSTL----- 908
 QY 324 SSGGYLTRHNGTTTPASVSIVLPBGFALERYDPNDGSWTDPAAGTQVTPROVAUDEV 383
 DB 909 -----NKYTPPTFNAT-----PDGNDMTAMYKEBIPFASVTLTRPKRV 947
 QY 384 VYINNPAAGGSAFTFVAVPPSNAYNTVFRNTLETPS-SRLELP--MPPADFGQTV 440
 DB 948 IDTNTNVYWTITITITISKAP-----LDNLTAKKPPNMSAGLTTFPMEVTEGETT 998
 QY 441 ANPKRISQSLKETLGCYLVS KKNPNVFLTPASSFGAVSP-----NNPGYERTRDLPD 495
 DB 999 KSIP-VNSTLWTEGV-----PLPNVPIGKIVSAFTBATGKPNVLKAEVVV 1046
 QY 436 YTGIRDSFDQNNSTVAHFRSLSHSCSIVTKYQMBEGYTNVNTFPQGAHAGLKNBEI 555
 DB 1047 FGGIRDSIVDNF-----VRIRPNDQEVVYTPTEBGFISVPTPD--FGVGVA GTQOQHS 1098
 QY 556 LCLAD-----DLATRLGVYPATDNF-----AAVASAPAN 586
 DB 1099 KQALDYNGNTRNPLRIKTPQPMNSLTQAQSPASATDSLPTATRLLGAAVPSFTYV 1158

QY 587 MLSSVLKSE--ATSSIIKSVGETA 608
 DB 1159 NQPTLKNVTGTTSAISLTANNTA 1182

RESULT 8
 AAY00218
 ID AAY00218 standard; Protein; 1265 AA.
 AC AAY00218;
 XX
 DT 20-APR-1999 (first entry)
 DE Enterococcus faecalis protein EF108.
 XX
 KM Enterococcus faecalis; infection; vaccine; immune response; diagnosis;
 KM detection; attenuation; antigenic.
 XX
 OS Enterococcus faecalis.
 XX
 PN WO9850554-A2.
 XX
 PD 12-NOV-1998.
 XX
 PF 04-MAY-1998; 98WO-US08959.
 XX
 PR 14-NOV-1997; 97US-0066009.
 PR 06-MAY-1997; 97US-0044031.
 PR 16-MAY-1997; 97US-0046655.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;
 DR WPI; 1999-070095/06.

DR N-PSDB; AAX20208.
 XX
 PT New isolated Enterococcus faecalis polynucleotides - used to develop
 PT products for the detection of Enterococcus and for use in vaccines
 PT for prevention or attenuation of Enterococcus infection
 PS
 XX Claim 9; Page 209-210; 301pp; English.
 XX
 CC The present sequence represents a protein isolated from
 CC Enterococcus faecalis. The present invention describes genes, proteins
 CC and antigenic polypeptides isolated from E. faecalis. The proteins can
 CC be used in vaccines for preventing or attenuating an infection caused
 CC by a member of the Enterococcus genus in an animal. They can also be
 CC used for detecting Enterococcus antibodies in a sample. The nucleotide
 CC sequences can be used for detecting Enterococcus nucleic acids.
 CC Products from the present invention can also be used for screening
 CC compounds to identify agonists and antagonists of E. faecalis protein
 CC activity.
 CC
 SQ Sequence 1265 AA;
 Query Match 4.1%; Score 139.5; DB 20; Length 1265;
 Best Local Similarity 19.0%; Pred. No. 0.036;
 Matches 141; Conservative 94; Mismatches 252; Indels 257; Gaps 31;
 QY 19 NVRVSAANTVTNNGRNRGRGQVSPDNFTAAADLAQSIDANTV---TFPANISS- 73
 DB 574 NRKVTENFVDTNGAR-----ITPTGFTGKKTVITS-DATYFQAAGTLPTTYTG 623
 QY 74 --MPEFRMAKGIKIDSDSIGWYFKYLDPAAGATSAVAGVSKIPDGLVFSVDABIR 131
 DB 624 GKTYFKGMVKOKSLINTLTIT-----TAPSYQVYDDNDILN 661
 QY 132 EYNEECVAVDVSVPLD-----GRQWSLSIFSPMERTAVAAVAVEN 175
 DB 662 VYEBETVTVYPSVDMNFVNEKGAFTPALTFSGKYVAQSTSAV--LRTDLYDVTSKNN 719
 QY 176 -----KMSLDVAVNDLIEMLN----- 191
 DB 720 GNGQYTVSINNQSMLSQELLKKNNGOPISATNRLQFNVDKLAIQDLKYVDSIQDLTA 779
 QY 192 ---NLADRRYVYV-----DSEQWINFNTDITYVY---RIRVLAR 223
 DB 780 QSSNLKSYRYVYTNSSSLVFPDPAVAEVDLSSESLINLINFSDGTYFSNANNRLPYTHL 839
 QY 224 TYD-VPDETEGLVRTVSD-----YRLTYKAITCEANMPTLVDOGFWIGQYALTP 272
 DB 840 GYSGTPGVNVLMLFNAKPADKSKLVYKTRKQVT-----ENFVD---VNGAKITAP 890
 QY 273 TSLPQYDVSEAVAAHTLTPARPSSAAL-----AFVWAGLPQGGTAPAGTAPAMEQA 323
 DB 891 TGFTD---GNOVPNMSNTF-KYTAAKALPATYTTGGKYVTFQGWYKKGKTKPSTL----- 940
 QY 324 SSGGYLTRHNGGTTPAGSVSVLPPEGALBERYDPDGSWTOFASAGDVTFRQYAVDVEV 383
 DB 941 -----NKTTFPTFRNAT-----FDGNDMTANYKSEIPLASTLTLRPEXV 979
 QY 384 VVTNNPAGGSAFTFTVAVPSNAYTNTVFRNTLTETRPS-SRRLELP--MPADPQGQTV 440
 DB 980 IDINTNVTWTTTITNTSKAP-----LQNLTLKQPNMSAGLTIPTFMEVTEGERT 1030
 QY 441 ANNPIEBSLKETIGCYLVHSKRKNPVQULTPASSFGAVSF-----NNQYERTRDLPD 495
 DB 1031 KSIPLVNSTLMTBEGV-----PLPNAVPICKKSVAVFTTRATGKPNVLAKAEVVV 1078
 QY 496 YTGIDSFQDNMSTAVAFPRSLSHSCSIVTKTYQGEVGTAVNTPPGQFAHAGLKNEBI 555
 DB 1079 FGGIDSTIVDNF-----VRIHPNDQEVVTPTEGFISSVPTD--FGQVGAAGTACQNSL 1130
 QY 556 LCLAD-----DLATRLGVVPAIDNF-----AAVSAFAN 586
 DB 1131 KQADYVNGTRNPLYRIKKTQPNWSLTAQISOPASATDSLPTATRLLLGAPVSSFTNY 1190

QY 587 MLSVLKSR--ATSSIKSVGETA 608
 DB 1191 NQPTLKNVTGTTSAISILNANTYA 1214
 RESULT 9
 ID ABP43437 standard; Protein; 1265 AA.
 XX
 AC ABP43437;
 XX
 DR 05-AUG-2002 (first entry)
 XX
 DE E faecalis EF108 protein.
 XX
 KM Enterococcus; vaccine; gastrointestinal disease; diagnosis; antibiotic.
 XX
 OS Enterococcus faecalis.
 XX
 PN US2002045737-A1.
 XX
 PD 18-APR-2002.
 XX
 PF 04-MAY-1998; 98US-0071035.
 XX
 PR 04-MAY-1998; 98US-0071035.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Choi GH, Bailey C, Hromochky A, Kunsch CA;
 DR WPI: 2002-425450/45.
 XX
 DR N-PSDB; ABR98193.
 PT New genes and polypeptides from Enterococcus faecalis, useful as
 PT vaccines for preventing, treating or attenuating an infection caused by
 PT a member of the Enterococcus genus in an animal, particularly E.
 PT faecalis -
 XX
 PS Claim 9; Page 193-194; 255pp; English.
 XX
 CC The present invention provides the protein and coding sequences of a
 CC number of polypeptides from Enterococcus faecalis. The proteins can be
 CC used as vaccines for preventing or attenuating an infection caused by a
 CC member of the Enterococcus genus in an animal, particularly E. faecalis.
 CC The polynucleotide is also useful for preventing or treating E. faecalis
 CC infection. The present sequence is a protein of the invention.
 CC
 SQ Sequence 1265 AA;
 Query Match 4.1%; Score 139.5; DB 23; Length 1265;
 Best Local Similarity 19.0%; Pred. No. 0.036;
 Matches 141; Conservative 94; Mismatches 252; Indels 257; Gaps 31;
 QY 19 NVRVSAANTVTNNGRNRGRGQVSPDNFTAAADLAQSIDANTV---TFPANISS- 73
 DB 574 NRKVTENFVDTNGAR-----ITPTGFTGKKTVITS-DATYFQAAGTLPTTYTG 623
 QY 74 --MPEFRMAKGIKIDSDSIGWYFKYLDPAAGATSAVAGVSKIPDGLVFSVDABIR 131
 DB 624 GKTYFKGMVKOKSLINTLTIT-----TAPSYQVYDDNDILN 661
 QY 132 EYNEECVAVDVSVPLD-----GRQWSLSIFSPMERTAVAAVAVEN 175
 DB 662 VYEBETVTVYPSVDMNFVNEKGAFTPALTFSGKYVAQSTSAV--LRTDLYDVTSKNN 719
 QY 176 -----KMSLDVAVNDLIEMLN----- 191
 DB 720 GNGQYTVSINNQSMLSQELLKKNNGOPISATNRLQFNVDKLAIQDLKYVDSIQDLTA 779
 QY 192 ---NLADRRYVYV-----DSEQWINFNTDITYVY---RIRVLAR 223
 DB 780 QSSNLKSYRYVYTNSSSLVFPDPAVAEVDLSSESLINLINFSDGTYFSNANNRLPYTHL 839

Qy	224	TYD-VPPBTECLVTVSP-----YRLTVKATICEAMPTLVDOGFQIGQVALTP	272
Db	840	GYSGETPGCVNYLLWFLFPAKPADSKVLKYKTRKQVT-----ENFVD-----VNAKITTAP	890
Qy	273	TSLPQDVSEAYALHTLTFARPSSAAAL-----AFVWAGLPQGGTADAGTPAMEQA	323
Db	891	TGFTQY-----GNQVPMNSNTF-KYTKAAKALPATYTTGAKYVTEQGWYKGTKEPSTL-----	940
Qy	324	SSGGVLTMRNHGTFPPAGSVSVYLEPEGALERYNDPNDGSMWDFASAGTVPFRQVANDV	383
Db	941	-----NKTTFPFNAL-----FQNDMTMYKBEIPIFASVTLTRPREV	979
Qy	384	VVTNNPAGGSGAPFTYVAVPPSNAYTNTVFANLTLETRPS-SRRLELP--MPPADFGQTV	440
Db	980	IDTNTNVIWTTIINTSKAP-----LQNLTLKKGPMSAGLTIPTFBHVPBEHETT	1030
Qy	441	ANNPKEIOSLKELTGCYLVSKKRNPVFOULTPSSFGASVF-----NNPGEYTRDLDP	495
Db	1031	KSIP-VNSTLTGEGV-----PLPNAVPPIGKKVSVAFPTTATGCKPMTVLKAEVVV	1078
Qy	496	YTGIRDSEPDQMSSTAVAHFRSLSHSCSVLVTYKQGWBEQSVTNVNPFGQFAPAGLLKNEBI	555
Db	1079	FGGJKDSTVDNF-----VRIRPDQEVVYTTTSGFISVPPFD-FQGVGAATGKQOHSL	1130
Qy	556	LCLAD-----DLATRLTGVPYPATDNF-----AAVASAFAN	586
Db	1131	KQAADVYNGTRNPFYLRIKKTQPMWSLTAQGSQPSKASDLSPTATRLLLGAAPVSSFTNY	1190
Qy	587	MLSSVLSKSE--ATSSIIKSVGETA	608
Db	1191	NOPELKNVGTGTSISILTANNTA	1214

RESULT 10
ABU13716
ID ABU13716 standard; Protein; 1265 AA.
XX
AC ABU13716;
XX
DT 26-FEB-2003 (first entry)
XX
DE Enterococcus faecalis EF040 polypeptide #209.
XX
KW EF040; immunostimulant; antibacterial; gene mapping.
XX
OS Enterococcus faecalis.
XX
PN US6448043-B1.
XX
PD 10-SEP-2002.
XX
PE 04-MAY-1998; 98US-0071035.
XX
PF 06-MAY-1997; 97US-044031P.
PR 16-MAY-1997; 97US-046655P.
PR 14-NOV-1997; 97US-066009P.
PR 14-NOV-1997; 97US-066009P.
XX
PA (HOMA-) HUMAN GENOME SCT INC.
XX
PI Choi GH, Bailey C, Hromockyj A, Kunsch CA;
XX
DR MPI: 2003-089120/08.
XX
DR N-P8DB; ABX61763.
XX
PT New EF040 polypeptides and polynucleotides from Enterococcus faecalis,
XX
PT useful for generating an immune response against B. faecalis and other
XX
PT Enterococcus species, and as vaccines against other bacterial genera
XX
PS Example 1, Column 215-216; 146pp; English.
XX
CC The invention relates to polynucleotide fragments of a gene from

CC Enterococcus faecalis, EP040, and the polypeptides encoded by them. The
CC polypeptides are useful in detecting *E. faecalis*, as epitope tags, as
CC molecular weight markers on SDS-PAGE gels or for molecular sieve gel
CC filtration columns, in generating antibodies that specifically bind to
CC the *E. faecalis* polypeptides, in generating an immune response against *E.*
CC faecalis and other Enterococcus species and as vaccines against other
CC bacterial genera. The polynucleotides are useful as probes for gene
CC mapping and for identifying *E. faecalis* in biological samples. Sequences
CC AB013508-AB013755 represent EP040 polypeptides of the invention.
CC Note: The sequence data for this patent can also be obtained from USPTO
CC at seqdata.uspto.gov/sequence.html.
XX
SQ Sequence 1265 AA;

Query Match	4.1%	Score 139.5;	DB 24;	Length 1265;
Best Local Similarity	19.0%;	Pred. No. 0.036;		
Matches 141;	Conservative 94;	Mismatches 252;	Indels 257;	Gaps 31

QY	19	NVR:SA:TVTVNGRRNRORRRTRGQVSPDPNFTAAQLOASLDANTV-----TFPANT:SS-73
Db	574	NKATKTFVDTNGAK-----ITPFGTQKQKTVITB-DATYTKQAGITLBDYTTTG623
QY	74	--MBEPFNMAKGKIDLDSDISGMYFKYLDPAGATESARAVGEXSKIDGLVKESVDABIR131
QY	132	EINHECPVTVDSVVPD-----GROMSLSIBFPMFRATYAVAVANVEN175
Db	662	VYBEETVTVTVYPPVDNMFPVNBKGAFTPALTPSGKYAOSTSAY--LRDLYDVTSKNN719
QY	176	-----KEMSDVANNLDLEMLN-----191
Db	720	GNQGYTVSINNGBSHPLOGBLLKTKNNQPI:SA:TRLRQPNVDKLAIDQGLKTVDSIQD:TA779
QY	192	---NLADMRVYV-----DSBQMINTNDTTYV---RIRVLAP223
Db	780	QSSNLKSYRYYYTNNSSLVFPEDPNVAPABAVDLSSESLIMLNFSDGTYFP:SNANMLFYTHL839
QY	224	TYD--VPDPTBELVXTVD-----YRLTYALICEANMPLNDQGFIMGQYALTP272
Db	840	GYSGTPEGVNYLLVNFLEFNAPADKSKLVYKTRKQVT-----ENFVD---VNAKITTAP890
QY	273	TSLPOYDVSBAVYLAHTLTPARPS:SAAL-----AFVWAGLPQCGTAPACTPAMEQA323
Db	891	TGTFQ---GNQVPMNSWTF-KYTA:KALBPATYTTGKGYTFQGYMKKGTKESTL-----940
QY	324	SSGGYLTWRHNGTTFPAGSVSYVLPEGPALERYPDNDGWTDFASAGDYTFTRQVAND:EV383
Db	941	-----NKTTTPTFNAT-----FGNDMDTMYKBI:PTASVTLTRPKEV979
QY	384	VVTNNPAGGSGAPFTYTRVPPSNAYTMTVPFNTLLSTRPS-SRLELP--MPPADPQGTV440
Db	980	IDTITNNVITWTTTINTSRAP-----LONLLTKGPMNSAGLITPTFMBVTPBGGTT1030
QY	441	ANNEKIBOSLKETLGCYLVH:SKARNVFPOLTLPAS:FGAVSF-----NNPGYERTRDLPD495
Db	1031	KSIP-VNSTLTBEV-----PLEPNAVPIGKKVSVAFITTRANGKPTVLKABVYV1078
QY	496	YTGIRDSFDQMS:TA:VAHAFSLSHS:CS:IVTKYTGMBEGVTNNVNTPPQGFPAHAGL:K:KEBI555
Db	1079	FGGJKIDSTVDNF-----VRIRPNDQRBVFTTGGFSVLPFD--FGQVGAGVAGKQOHS:L1130
QY	556	LC:LA:ND-----DLATRLTGYVPATDNE-----AAASAFAPAN586
Db	1131	KQADYDYGNGTRNRYLRIKKTQPMWSLTPAOLDSQPSK:ATD:SLPTATRLLLGHAAPVSSFTNY1190
QY	587	ML:SVL:KSE--AT:SI:IK:SVGE:TA608
Db	1191	NQPTELKNVTGVTSAISILTANNTA1214

ID	AB866424 standard; Protein; 2016 AA.
XX	
AC	AB866424;
DT	
XX	
DE	26-MAR-2002 (first entry)
XX	
DE	Drosophila melanogaster polypeptide SEQ ID NO 26064.
XX	
KW	Drosophila: developmental biology; cell signalling; insecticide;
XX	
OS	pharmaceutical.
XX	
OS	Drosophila melanogaster.
XX	
PN	WO200171042-A2.
XX	
PD	
XX	
XX	27-SEP-2001.
XX	
PE	23-MAR-2001; 2001WO-US09231.
PR	23-MAR-2000; 2000US-191637P.
PR	11-JUL-2000; 2000US-061415O.
XX	
PA	(PEKE) PE CORP NY.
XX	
P1	Venter JC, Adams M, Li PWD, Myers EW;
XX	
DR	WPI: 2001-6556860/75.
DR	N-PSTDB; ABLI0527.
XX	
PT	New isolated nucleic acid detection reagent for detecting 1000 or more genes from Drosophila and for elucidating cell signalling and cell-cell interactions -
XX	
PS	Disclosure; SEQ ID NO 26064; 21pp + Sequence listing; English.
CC	The invention relates to an isolated nucleic acid detection reagent capable of detecting 1000 or more genes from Drosophila. The invention is useful in developmental biology and in elucidating cell signalling and cell-cell interactions in higher eukaryotes for the development of insecticides, therapeutics and pharmaceutical drugs. The invention discloses genomic DNA sequences (ABLI01840-ABLI6175) and the encoded proteins (ABB57737-ABB72072).
CC	sequences (ABLI01840-ABLI6175) and the encoded proteins
CC	(ABB57737-ABB72072).
CC	The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
CC	
XX	
XX	Sequence 2016 AA;
SQ	
Query Match	3.8%; Score 127; DB 22; Length 2016;
Best Local Similarity	20.7%; Pred. No. 0.81;
Matches	77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;
Db	
Oy	231 TEGLRVTSYRLTYKAITCEANMPTLVDOGFMI-----GGQYALPTSIPDYVSEA 283 ::: :::
Db	1259 TETKTQKVPHQMSEATELEKNKP---YEPWTASTTIGGQQSKSIVAMPSPQVPAK 1314 ::: :::
Oy	284 YA----LHTLPARRSSAAALAFAVAGLPQGCTAAGAPNMGASGGYLTRKHGTTFP 339 ::: :::
Db	1315 IASPDFTFATFKEDAKMPCLA--VGAPQ-----BEITWKIKGVFS 1354 ::: :::
Oy	340 AGSVSYVLPEGFAL-----ERYDPNDGS-WTPFASAGDVTFRQAVN----- 380 ::: :::
Db	1355 ANDMRNVLPDSLSILKSIVNRQDAGVSGCHAENSIAKSDITHKLIVLARPPQSRHVLSAT 1414 ::: :::
Oy	381 -DEVVVTTNNPAGGSAPT--FTVRVPQS-----NAYTN 410 ::: :::
Db	1415 TDALTVLKKPHBGDPAPLHGTYLHKPFEGEWETSVEVDSQKHIEGLCGSRQYYVA 1472 ::: :::
Oy	411 TVFN-----TLLETRPSSRRLELPMRPADFGQTVANNPRIEBSLKLKETLCG---YL 459 ::: :::
Db	1474 TGFNNGICGEASDLINRTTKGPKPLPEKPR-FLEVSSNSVSLHFKAAMDG-GCPMSHPV 1531 ::: :::

QY	460	VHSHKRN-----VQLTASSFG-AVSPNNPGY-----	487
DB	1532	VBSKKRQDIEWNISSNVKPDNNVYVLDLBPATWYNIIRITAHNSAGFTVAEYDPATLTVT	1591
QY	488	----BRTDLPD 495	
DB	1592	GGTIAPSRDLPE 1603	
RESULT 12			
ABG73274	ABG73274 standard; protein, 2016 AA.		
AC	ABG73274;		
XX			
DT	22-APR-2003 (first entry)		
XX			
DE	D. melanogaster Down syndrome cell adhesion molecule (DSCAM) protein.		
XX			
KM	RNA alignment coverage; sparse alignment coverage; extension;		
KM	RNA derived sequence fragment; catenation; overlapping alignment;		
XX	exon; Down syndrome cell adhesion molecule; DSCAM.		
OS	Drosophila melanogaster.		
XX			
PN	US2002150895-A1.		
PD	17-OCT-2002.		
XX			
PF	22-DEC-2000; 2000US-0747440.		
XX			
PR	22-DEC-2000; 2000US-0747440.		
XX			
PA	(WHBE/) WHEELER R.		
XX			
PI	Wheeler R;		
DR	WPI; 2003-182492/18.		
XX			
PT	Extending sparse alignment coverage for any combination of RNA derived		
PT	sequence fragments by combining and catenating all combinations of		
PT	overlapping alignments, and extending boundaries of overlapping		
XX	alignments -		
PS	Disclosure; Page 7; 35pp; English.		
XX			
CC	The present invention relates to a method for filtering and extending		
CC	RNA alignment coverage. Sparse alignment coverage for any combination		
CC	of RNA derived sequence fragments are extended by combining and		
CC	catenating all combinations of overlapping alignments that agree with		
CC	each other, and extending boundaries of overlapping alignments that		
CC	agree with their first and last exons. The method is useful for		
CC	extending sparse alignment coverage for any combination of RNA		
CC	derived sequence fragments. The invention is capable of extending		
CC	a typically sparse RNA alignment coverage, without creating redundant		
CC	or improbable alignments. The present sequence representing Drosophila		
CC	melanogaster (Down syndrome cell adhesion molecule) DSCAM protein		
CC	is described in the specification of the present invention.		
XX			
SQ	Sequence 2016 AA;		
Query Match	3.8%; Score 127; DB 24; Length 2016;		
Best local similarity	20.7%; Pred. No. 0.81; Mismatches 116; Indels 134; Gaps 18;		
Matches	77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;		
QY	231 TEGLVRTVSDYRLTYKALITCEANPPTLVDOGFWI-----GGQYALTPSLPDYVSEA 283		
DB	1259 TETTKOKPVRXQMSFATELEKNP-----YEFVTAATGIBGGQSSIVAMPDQVPAK 1314		
QY	284 VA-----LHTLPAPSSAALAPFVAGLPGCGTAPAGTPAMRQASSGGYLTWRANGTTP 339		
DB	1315 IASPDPTTATKEDAKMPCIA---VAAPQ-----PEITWKIKGVFFS 1354		

```

QY 340 AGSVSYVLPBEPAL-----ERYDPNDGS-WTFASAGDTVTRQAVN----- 380
DB 1355 ANDRMKVLPGDGLLKSVNRQDAGYSCHAENSIMADSIHTKLIVLPQSPHTLSATT 1414
QY 381 -DEVVVTNNPAGGSAPT--FTVRVPS-----NAYNN 410
DB 1415 TDLATVTKLPHEGDTAPLHGYTLHYKPEGEMETSEVSDQKHNIGLLCGSRQVYA- 1473
QY 411 TVFRN-----TLLETRPSRRLLEPMPADFGQVANNPKIBOSLKEITLGC-----YL 459
DB 1474 TGFNNIGAGEASDIINTRKQKPKLPEKPR-FIEVSSNSVSLHPRAMWGQ-GCPMSHPTV 1531
QY 460 VHSKRNPP-----VFQUTPASSFG--AVSFNNPGY----- 487
DB 1532 VESKRRDQIEMNQISNNVAPDNNVYVLDLEPATWYLRITAHNSAGFTVAEYDFATLVT 1591
QY 488 -----ERTRLPD 495
DB 1592 GGTIAPSRLDPS 1603

```

RESULT 13

ABP43908
ID ABP43908 standard; Protein; 4315 AA.

XX ABP43908;

DT 26-FEB-2003 (first entry)

DE MUC5B partial gene protein.

XX Neutrotective; immunomodulator; cancer; chromosome 11p15;

KW cytostatic; anti-inflammatory; gene therapy; nutritional supplement;

KW wound; burn; ulcer; Alzheimer's disease; Huntington's disease;

KW amyotrophic lateral sclerosis; autoimmune disorder; inflammation;

XX vulnery.

OS Homo sapiens.

XX MO200231111-A2.

PD 18-APR-2002.

PF 11-OCT-2001; 2001WO-US27760.

PR 12-OCT-2000; 2000US-0687527.

PA (HYSE-) HYSEQ INC.

PI Tang YT, Liu C, Zhou P, Asundi V, Zhang J, Zhao QA, Ren F;

PI Xue AJ, Yang Y, Wehrman T, Drmanac RT;

XX WPI; 2002-426278/45.

DR N-PSDB; ABQ61152.

XX New polypeptides and their encoded proteins, useful as nutritional

PT sources or supplements, or in gene therapy, particularly for treating

PT wounds, Alzheimer's disease, amyotrophic lateral sclerosis, cancer or

PT inflammation

XX Claim 20; SEQ ID # 811; 357pp + sequence listing; English.

CC The invention relates to 446 newly isolated polynucleotide sequences.

CC The activity of polynucleotides of the invention may be described as,

CC vulnery, neutrotective, immunomodulator, cyostatic and

CC anti-inflammatory. Compositions comprising nucleic acids of the invention

CC are useful for treating a mammalian subject, or as nutritional sources or

CC supplements. These are useful in gene therapy, particularly for treating

CC wounds, burns or ulcers, Alzheimer's disease, Huntington's disease,

CC amyotrophic lateral sclerosis, autoimmune disorders, cancer or

CC inflammation. The nucleic acids and polypeptides are also useful in

CC diagnostic and research methods. The sequences given in records in

CC ABP43544-ABP43589 represent polypeptides encoded by polynucleotides of

CC the invention.

CC NOTE: The sequence data for this patent did not form part of the printed

CC specification, but was obtained in electronic format directly from WIPO

CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 4315 AA;

Query Match 3.7%; Score 126; DB 23; Length 4315;

Best Local Similarity 22.3%; Pred. No. 3.1;

Matches 94; Conservative 32; Mismatches 178; Indels 118; Gaps 14;

QY 221 LRPTVDVDPBEGVLRVTSRRLTY---KAITCANPFLVDGPFMIGGQALPTSLP 276

DB 580 LNSTATPTTATSPFAIPSSISGTTWRLSQTTTPATNST-----ATPSSSTP 626

QY 277 QYDVEAVALHTLT-----FAPSSAALAPFWAGLPGGTAPAGPAMEQ 322

DB 627 E-----TVHTSTVLTATTAATGAGVATPSSIPGTNAHTTKVLTITTTGPTATPSSSP 679

QY 333 ASSGGYLTWRHNGITTPAGSVSYVLPBEPALERYDPNDGSWTFASAGDTVTRQAVDE 382

DB 680 GRATLPVW-ISTTTPTTSGSTVTPSSI-----PGTHPTVLTITTT 721

QY 363 VVYTNPPAGGGSAPFTVTRVPPSNAYTNTVTRNTLETRPSSRLLEPMPADFGQVANN 442

DB 722 TVVATGSMATPSSSTQTSQTPPSLTATTAATTAATGAGVATPSSIPGTNAHTTKVLTITTTGPTATPSSSP 779

QY 443 NPKIBOSLKEITLGCYLVHKKRNPPVQUTPASSFGAVSFPNNPGYERTDLPDYGIRDS 502

DB 780 TPAATSS-----TVTPSSALGTT-----HTPPVNN----- 804

QY 503 PDQNNSTAVAFRSLSHS-----CSIVTKYQMGVNTVNTPFQGFAN-----AGLKN 552

DB 805 -----TATTHGRSLSPSPHTVCAWTSATSGILGTHITPSTGSHTPAATGTTQH 859

QY 553 EETLCLADDLATRLTGTVPA-----TDNFAAVSAPAA-----NKLSSVLSKATSSITLK 602

DB 860 STPALSPHPSRTSPSPSPPTGHTTATSRITATATPASKRTSTLPSQPTSPAPT 919

QY 603 SV 604

DB 920 TV 921

RESULT 14

ABP75861

ID ABP75861 standard; Protein; 1093 AA.

XX ABP75861;

DT 10-FEB-2003 (first entry)

DE Human secretory polypeptide SPTM SEQ ID NO 1045.

XX Human, SPTM; autoimmune disorder; inflammatory disorder; AIDS; anaemia;

KW asthma; Crohn's disease; neurological disorder; epilepsy; cancer;

KW Huntington's disease; Alzheimer's disease; Creutzfeldt-Jakob disease;

KW anti-inflammatory; Parkinson's disease; cell proliferative disorder;

KW neuroleptic; anticonvulsant; cyostatic; antiparkinsonian; anxiolytic;

KW secretory polynucleotide; secretory protein.

XX Homo sapiens.

PN WO200283876-A2.

XX 24-OCT-2002.

XX 27-MAR-2002; 2002WO-US09921.

PR 29-MAR-2001; 2001US-280067P.

PR 29-MAR-2001; 2001US-280068P.

CC antibodies, identification of L. monocytogenes and related organisms, and
CC for biosynthesis and biodegradation, especially biosynthesis of Vitamin
CC B12. The genome sequence and proteins encoded by it are also useful for
CC selecting compounds that regulate gene expression and cell replication
CC and modulate L. monocytogenes-related diseases. In addition, the genome
CC sequence and proteins encoded by it are useful in pharmaceutical and
CC vaccines compositions for the treatment or prevention of infections by L.
CC monocytogenes and related organisms.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 1582 AA;

Query Match 3.6%; Score 122; DB 23; Length 1582;

Best Local Similarity 19.6%; Pred. No. 1.5;

Matches 122; Conservative 92; Mismatches 247; Indels 160; Gaps 35;

QY 31 GRRNRRRTGRQVSPDNTAAQDLAQSIDANTVTPPANISSMPEFRMAKGIIDSD 90
DB 31 GABENQTKTQQAQDDQKVADEN-----KTVVTPNKEIT---KNLVNKG-QLKG 79
QY 91 SIGWVF--KYLDPAATESARAVGEYSKI PDGLVKFSV-----DAEIRBIYN----- 135
DB 80 SYDMNFTNKVTDANNGTKETYPAD-----SLKFSVLAIPTSYSEAVLDEITYTFWIOK 132
QY 136 --EECPVVTDVSVPLDGRQWSLISFSPMERFA-----YV-----AVAN 172
DB 133 DFGQGDKVTFFVNPALIDGLPRQNVFTDDSMKNSIGTKINGVEYIGITLKFNSPEALAK 192
QY 173 VENKENS---LDVNDLIEWLNADMR-YVDSQMINFTDITYVRIRVLRPTVDV 228
DB 193 YEDPEPLTNKFPDMFATISTITTKEDYKAIVIEDDDVDISN-----IEVPTVPT---P 244
QY 229 DPREGLVRTVSDVRLTYKAITCEANMPTLVDOGFWIG--GOYALPTSLEPOYVSEAVA 285
DB 245 DTDNG-VLTAS-----KVIDSAMRLDKDTGKIVKATISSTNTTPRKDL 287
QY 286 LHTLTFARPSSA--AALAFVWAGLPOG-GTAPAGTPAMEQAS--SGGYLT--W--RHNGT 336
DB 288 VFYYVYAMNNSGENAVLEKIMIDELPAQFSITTEGSPWQALSTASGSSLORGWVASADQT 347
QY 337 TPPAGSVSYLPEGFALERIDPDGSGTDPASAGDVTTPQVAV-----DEVVYTNMP 389
DB 348 NLAPGATRYELPSPSA-QTISKNGG-----IYRTIAKYTDPTKDLTNTYKSP 395
QY 390 AGGSAPTFTVRVPSPNAYNTVFRNTLLETSPSRRLLELPMPPADFGQTVANPKIEQS 449
DB 396 TKGGEGKSGSITPSGDVY-----DMAITTKISS-----TYDKNKY--- 432
QY 450 LKETLGCYLVHSGKMRNPVQLTPASSFGA--VSFNNPGYERTDL PDYTGIRDSFDQM 507
DB 433 -----IGAV---SSTTPV-SISEGSTVGVTYAINQNY--TKDVVYVAYVPAGEELNN 481
QY 508 STAVAFRSLSHCSIVTKTYQG---WEGVTNNTPPGQFAHAGLKNBEILCLADDIAT 564
DB 482 DT--SDFKTXANDKWKETVPSGGVWSIDIKVSYITLSGGMASGATKTATMYKA--MGI 537
QY 565 RLTVVYPATDNF-AAAVSAPA 584
DB 538 PEDGTYSADVFMAGEIGSFS 558

Search completed: January 15, 2004, 16:49:14
Job time : 26.9426 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 16:26:38 ; Search time 12.9345 Seconds

(without alignments)
4810.480 Million cell updates/sec

Title: US-09-991-262-50

Perfect score: 3374

Sequence: 1 MGDAVVASQPHNRGRTRNV.....GKTAARVARRARRARRAN 647

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2179	64.6	644	2 A43370	capsid protein - N
2	130	3.9	597	2 B75556	hypothetical prote
3	130	3.9	708	1 OQBER8	Urb0 protein - hum
4	126	3.7	3570	2 T45025	mucin MUC5B, trach
5	125.5	3.7	1269	2 A90267	proteinase related
6	125	3.7	2314	2 T28698	hypothetical prote
7	123.5	3.7	1204	2 C75015	probable pyrolysin
8	123	3.6	886	2 S29605	glycoprotein 350/2
9	122	3.6	1582	2 AC1153	adhesin homolog lm
10	121.5	3.6	907	1 OQBER2	membrane antigen 5
11	120	3.6	873	2 B75514	penicillin-binding
12	119.5	3.5	889	2 T09055	protocadherin 68 -
13	119.5	3.5	2271	2 F90073	hypothetical prote
14	119	3.5	1363	2 T43220	insulin-like growth
15	119	3.5	1993	2 AF1450	probable peptidogl
16	119	3.5	2809	2 T30213	G-cadherin - sea u
17	118.5	3.5	3029	2 S76109	hypothetical prote
18	117.5	3.5	609	2 S62518	conserved hypotet
19	117.5	3.5	884	2 B75489	conserved hypotet
20	117.5	3.5	26926	1 T38344	titin, cardiac mus
21	117	3.5	674	2 A97480	hypothetical prote
22	117	3.5	859	2 AE2217	hypothetical prote
23	116	3.4	3716	2 E70969	probable PPS prote
24	115.5	3.4	3624	2 AD0835	chitinase (BC 3.2.
25	115.5	3.4	3624	2 AD0835	large repetitive p
26	115	3.4	1419	2 T30531	agglutinin-like ad
27	114.5	3.4	710	1 OQBER2	membrane antigen 5
28	114	3.4	828	2 T12184	probable transcrip
29	113.5	3.4	781	1 VCPV19	coat protein VP1 -

30	113	3.3	4436	2 B71086	hypothetical prote
31	112.5	3.3	604	2 A84753	hypothetical prote
32	112	3.3	1090	2 S59077	cellulose 1,4-beta
33	112	3.3	1578	2 AD1512	peptidoglycan bou
34	111.5	3.3	604	2 T31042	hypothetical prote
35	111	3.3	1902	2 B45764	lactocarpin (BC 3.4
36	110.5	3.3	505	2 S39962	endoglucanase - Br
37	110.5	3.3	677	2 B75563	probable peptidyl-
38	109.5	3.2	792	2 F88656	protein p56d6.1 (1
39	109.5	3.2	1662	2 T18540	mo6a protein precu
40	109	3.2	591	2 B87361	flagellar hook pro
41	109	3.2	766	2 B86597	C711 hypothetical
42	109	3.2	766	2 F72027	conserved hypotet
43	109	3.2	1302	2 T30191	multidrug resistan
44	108.5	3.2	642	2 B72428	laminarinase - The
45	108	3.2	424	2 T32434	hypothetical prote

ALIGNMENTS

RESULT 1	A43370	capsid protein - Nudarelia omega virus
C/Species:	Nudarelia omega virus	
C/Date:	16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 24-Nov-1999	
C/Accession:	A43370	
R/Accession:	D.K.; Johnson, J.E.	
Virology 190,	806-814, 1992	
A/Title:	Sequence and analysis of the capsid protein of Nudarelia capensis omega virus	
A/Reference number:	A43370; MUID:92391097; PMID:1519360	
A/Accession:	A43370	
A/Status:	preliminary	
A/Molecule type:	mRNA	
A/Residues:	1-644 <AGR>	
A/Cross-references:	GB:843937	
C/Superfamily:	Nudarelia omega virus capsid protein	
Query Match	64.6%; Score 2179; DB 2; Length 644;	
Best Local Similarity	64.7%; Pred. No. 2,3e-145;	
Matches	425; Conservative 68; Mismatches 130; Indels 34; Gaps 7;	
QY	8 SQRPHNRGRTRNV-----SANTVTVNGRNRQRRRGVSPPNFTAAADLA 57	
DB	3 SNSASGRRRSRNRIARQARRRARSANTVNVAPK-----QRANNINDVTAAGELG 54	
QY	58 OSLDANTVTPRAVSSMPEFRNAKGIKIDDSIGFPTPLDPAGATTSARAVGERSKI 117	
DB	55 OSLDANVTTPTNVATPBRFSMARGLDIDDSIGFPTPLDPAGATTSARAVGERSKI 114	
QY	118 PDGLVKEVDAREIREINRESCPVTVDSVPLDGRQMSLISFPMFPTAVAVANVENSKE 177	
DB	115 PDGLVKEVDAREIREINRESCPVTVDSVPLDGRQMSLISFPMFPTAVAVANVENSKE 174	
QY	178 MSLDVVDNLIEMNLNLDMDRYVVDSEQVINFNTDTTYVAVIRVLRPTVDVDPTEGLVRT 237	
DB	175 ISLDVTNDLIEMNLNLDMDRYVVDSEQVINFNTDTTYVAVIRVLRPTVDVDPTEGLVRT 234	
QY	238 VSDVRLTYKATICEANNPVTVDOGFWIGGOYALTFSTLPOYDVSBAVALHTLTARPSA 297	
DB	235 CSDFRLTYKATICEANNPVTVDOGFWIGGOYALTFSTLPOYDVSBAVALHTLTARPSA 294	
QY	298 AALAFVAGDPGCGTAPAGTPAMQASGGVLTWRNGTTPAGSVVYVLPBGALRYD 357	
DB	295 AGVTLTWASMPGGSAPSGDPAMIPDSITGP-QWRHGGPAPAPGVITTYITPRGTMYD 353	
QY	358 PNDGWTDPASAGDVTTFRCVAVDEVVTVNPPAGG-----SAPFTVRV-PPSNAYTNT 411	
DB	354 TTTNENNGFANPDVDVTFQ-----TGAAGTNAITITTAFTVLTTLIATTSANV 405	
QY	412 V-PRNTLLRTPSSRLLEMPAPDRCQTAAVNPKHOSLTKETLGCYLVHSKRNVPQ 470	
DB	406 INFRNLDAETTAASNSREVLPLPLTFEQTAPNNPKIQTIVKDTLGSYLVHSKRNVPQ 465	

```

Qy 471 LTPASSFGAVSNPNFYERTRLDPVYTGIRDSFDONMSTAVAFPSLSHSCSYVTKYQG 530
Db 466 LTPASSFGAISTPNQFDBNLDPGPGGIRDSLDMVMSYAVCHPSLSKSCSYVTKYQG 525
Qy 531 MEGVTNNVTPFGQFAHAGLLKNEBILCLADDLATRLTGVPYATNDPAAVSAFPAANMLSS 590
Db 526 MEGVTNNVTPFGQFAHAGLLKNEBILCLADDLATRLTGVPYATNDPAAVSAFPAANMLSS 585
Qy 591 VKSEATSIISVETAVGAAGSLAKLPGLLMSVPGKIAARVBARBARBARABAN 647
Db 586 VKSEATSIISVETAVGAAGSLAKLPGLLMSVPGKIAARVBARBARBARABAN 642

```

RESULT 2

```

B75556
Hypothetical protein - Deinococcus radiodurans (strain R1)
C.Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
C.Species: Deinococcus radiodurans
C.Accession: B75556
R.White, O.; Eichen, J.A.; Heidelberg, J.F.; Hickey, B.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uitterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A.Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A.Reference number: A75250; MUID:20036896; PMID:10567266
A.Accession: B75556
A.Status: preliminary
A.Molecule type: DNA
A.Residues: 1-597 <WHI>
A.Cross-references: GB:AE001876; GB:AE000513; NID:g6457800; PIDN:AAFO9732.1; PID:g645780
A.Experimental source: strain R1
C.Genetics:
A.Gene: DR0143
A.Map position: 1

```

```

Query Match 3.9%; Score 130; DB 2; Length 597;
Best Local Similarity 20.6%; Pred. No. 0.36;
Matches 128; Conservative 67; Mismatches 253; Indels 172; Gaps 25;

```

```

Qy 54 ODLAGSLDANTVTPANISMSPEFR---NMAKGIKIDSDSIGWFKYLDPAAGAT---- 105
Db 4 RIVDSLPAAGGGQGVVSTSPYQAVCVNNGSGIISNNSGNSLSYSSSDSGATVPPA 63
Qy 106 ---ESARAVGEYSKIPDGLVVFSDAIRE--IYNECP--VTVDSVPLDGRQWSLSI 157
Db 64 PTAGATQMSVPGNTIIPATCFRFSFVTLPRAGTYGNVLPANIRISGNDADGAGASTIA 123
Qy 158 FSPFPRFRTAY-----VAVANENKEMSLDV-VNDLIEMLANLADMRVYV 200
Db 124 FGPPTVTAKVSPKSLGDGKATSTLTITLNNSSPAVSLTAPLTDNIG--NGLBITGVYT 181
Qy 201 DSEQMINFNDDTYVYRIRVLRP-----TYDVPDPEGLVTRVSDVRLYYKAITCBANP 255
Db 182 SCGPATRTSGTITTYPSGATLMPGCTTITATYRSATAG-----STP 222
Qy 256 TLVDQG---FWIGQGYA---LTPSLPOYVNSEAYA-----LHTLTLPAPSSA 297
Db 223 NTISAGALQTTVGNNAASADLTLYTSFRLTITKTHASQNFPAQGTGYTLYTVAASGA 282
Qy 298 AALAVVMAGLPQGGTAPACTPAMEQASGGYLTWRHNGTTPAGSVSYLPEGFALERTD 357
Db 283 AA-----TSQALSLSDLLPSGMSFNSTV 305
Qy 358 PN-DGSMWDPASAGDVTFRQVAVDEVVVTNNPAGGSAFTFVVRPSPNATNTVFRMT 416
Db 306 TTAGGSFGTRPASGAT-----GRVDWTFPSTPLAAGSLTITVYVNVANVYANGATLTN 360
Qy 417 LLETSPSSRRLELPMPPAD-FQQTVAANNPKIEQSLKETLGCYLVHSHKRNKRVFQLT-- 472
Db 361 YASVGGGDPDVLFPFGATCTGEGCASDFTV-----NMITQTLAK 402
Qy 473 ---PASSFGAVSNPNFYERTRLDPVYTGIRDSFDONMST-AVAHPSLSHSCSYVTKYQ 529

```

```

Db 403 EPPQGAAPKPKFAN-----YDYACT--TIGGSSSTLTATTATSSIANPQVATDT-- 449
Qy 530 MEGVTNNVTPFGQFAHAGLLKNEBILCLADDLATRLTGVPYATNDPAAVSAFPAANML 587
Db 450 -----INITPGANYTLARLKDNDVAVSGFDSYDSR-----YTCTNATGSGTIVNPTSS 498
Qy 588 -LSVYLSKSEATSIISVYGR 606
Db 499 GMSFTLTTPQAGDIITCSVGD 518

```

RESULT 3

QGBB8

```

UL80 protein - human cytomegalovirus (strain AD169)
N.Contains: capsid assembly protein, viral proteinase (EC 3.4.21.-)
C.Species: human cytomegalovirus, human herpesvirus 5
A.Note: host Homo sapiens (man)
C.Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 16-Jul-1999
C.Accession: S09843; S51034; S51035
R.Chee, M.S.; Bankier, A.T.; Beck, S.; Bohnl, R.; Brown, C.M.; Cerny, R.; Hornell, T.
M.; Barrell, B.G.
Curr. Top. Microbiol. Immunol. 154, 125-169, 1990
A.Title: Analysis of the protein-coding content of the sequence of human cytomegalovir
A.Reference number: S09743; MUID:90269039; PMID:2161319
A.Accession: S09843
A.Status: nucleic acid sequence not shown; translation not shown
A.Molecule type: DNA
A.Residues: 1-708 <CHB>
A.Cross-references: EMBL:X17403; NID:g59591; PIDN:CAA3535.1; PID:g1780857
A.Note: possible protein-coding frames are given
A.Note: the DNA sequence was submitted to EMBL, December 1989, in computer-readable fo
R.Stevens, J.T.; Mapeili, C.; Rao, J.; Hall, M.; O'Boyle II, D.; Weinheimer, S.P.; Di
Bur, J. Biochem. 226, 361-367, 1994
A.Title: In vitro proteolytic activity and active-site identification of the human cyc
A.Reference number: S51034; MUID:95094793; PMID:8001553
A.Accession: S51034
A.Molecule type: protein
A.Residues: 110-131, 'X', 133-134 <ST2>
A.Note: assembly protein release site (A1a-256-Ser-257) and maturation site (A1a-643-S
C.Superfamily: cytomegalovirus capsid assembly protein
C.Keywords: capsid assembly, hydrolyase, serine proteinase
F.1-256/Product: viral proteinase #status predicted <PMAT>
F.136-708/Product: capsid assembly protein #status predicted <CAP>
F.132/Active site: Ser #status experimental
F.256-257/Cleavage site: Ala-Ser (viral proteinase) #status experimental
F.643-644/Cleavage site: Ala-Ser (viral proteinase) #status experimental

```

```

Query Match 3.9%; Score 130; DB 1; Length 708;
Best Local Similarity 21.9%; Pred. No. 0.47;
Matches 89; Conservative 42; Mismatches 151; Indels 124; Gaps 15;

```

```

Qy 269 ALTPSLPOYVSEAYAL-HTLTLPAPSSAALAFVWAGLPQGGTAPAGTAMEQASGG 327
Db 319 AESPPLSPSEBPAASAKSHPLSAVPAATA-----DPGATVAGASPA---VSS-- 364
Qy 328 YLTWRHNGTTPAGSVSYLPEGFALERTDYPNDGSMWDPASAGDVTFRQVAVDEVVYTN 387
Db 365 -LAWPHDQVLYPKQAFSL----- 383
Qy 388 NPAGGSAFTFVVRPSPNATNTVFRNTLTETSPSSRLBLPM--PADPQQTVAANNPK 445
Db 384 ----GASHSAVFVMYFGVVA-----APPSAPAPLPPLPSYASVYGAFAVGYDQ 427
Qy 446 IEQSLKETL-----GCYLVHSHKRNKRVFQLTTPASFGAVSNPNFYERTRLDP-- 495
Db 428 LAARFAYVDYDPHYGKRRYRPAASLPSFVPPPS-----PATYRRRDSFGGMD 479
Qy 496 -----YTGIRDSFDONM-----STAVAHPSLSHSCSYVTKYQGG 533
Db 480 EPPSGMERVYDGGHRQSQKQHRHGGSGGHNKRRKETAASSSSDSDLDSPQEAHGBAR 539
Qy 534 ---VTNNVTPFGQFAHAGLLKNEBILCLADDLATRLTGVPYATNDPAAVSAFPAANMLSS 590

```

Db 540 KRUKSHVNSDGGSGGAG--SNQOQQRVDELDAIHEL--KIDLFAARQS---STLLSA 592
 QY 591 VLKSEATSSIIKVGRTAVGAASGLAKLPGLMSVSGKTAAYRA 636
 Db 593 ALPSAASSSTTTTCTPTELTSGGETPTALLSGAKVAERQA 638

RESULT 4

T45025
 mucin MUCB, tracheobronchial [imported] - human (fragment)

C/Species: Homo sapiens (man)
 C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
 C/Accession: T45025
 R/Deasyam, J.L.; Guyonnet-Duperrat, V.; Porchet, N.; Aubert, J.P.; Laine, A.
 J. Biol. Chem. 272, 3168-3178, 1997
 A/Title: Human mucin gene MUC5B, the 10.7 kb large central exon encodes various alternat
 A/Reference number: 222899; MUID:9716151; PMID:9013550
 A/Accession: T45025
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-3570 <DES>
 A/Cross-references: EMBL:Z72496; NID:g1834502; PIDN:CAA96577.1; PID:g1834503
 A/Experimental source: placenta
 C/Genetics:
 A/Genes: MUC5B

Query Match 3.7%; Score 126; DB 2; Length 3570;
 Best Local Similarity 22.3%; Pred. No. 11;
 Matches 94; Conservative 32; Mismatches 178; Indels 118; Gaps 14;
 QY 221 LRPTVVPDPTGELVRTVSQYRLTY-----KATCEANMPTLVDOGFWIGQVALPTSLP 276
 Db 643 LRRTATTPATSTALPSSSLGTTWRLSGTTTPMAIMST-----ATBETSP 669
 QY 277 QYDVSAVALHTLT-----FARPSAALAFVWAGLPQGTAPAGTPAMEQ 322
 Db 690 E-----TVHTSTVLTTATTTGATGVSATPSTPCTAHTTKVLTGTTGATPSSP 742
 QY 323 ASSGGVLTMRHNGTTPAGSVSYLPGFALERYDPDGSWTBPASAGDTVTROQVAVD 362
 Db 743 GRARLTPVW-ISTYTTPTTGTSTVPSSI-----PGTHPTVLTLLTT 784
 QY 383 VVVTTNPPAGGSAPTFVVRPSPNAVNTVFRNLTLETRPSRLLEPMPAPFGQIVAN 442
 Db 785 TVVATGSMATPSSSTGTSTPSTLTATTTATGTTNPSSTPSTPPIPV--LTTAT 842
 QY 443 NPKIEOSLKEITGLCVLVHSMKNRPVQLTPASSFGAVSFNNPGYERTDLPDYTGIRDS 502
 Db 843 TPAATSS-----TVTPSSALGTT-----HTPPVPN----- 867
 QY 503 FPDQMTSTAVAHFSLSHS-----CSIYTKTYQMEGTNNVNTPPGQPAH-----AGILKN 552
 Db 868 -----TTATTHGRSLSPSSPHVTCTMTSATSGILGTHTEPSTGSHTPAATTTGTH 922
 QY 553 BEILCLADLATRLTGVPYPA-----TONFAAASFAA-----NMLSSVKSATSSIIK 602
 Db 923 STPALSSPHSPSSKTTSPSPGTTTTPGHTTATRTATPSTKRTSTLPSQTSAPIT 982
 QY 603 SV 604
 Db 983 TV 984

RESULT 5

A90267
 protease related protein [imported] - Sulfolobus solfataricus
 C/Species: Sulfolobus solfataricus
 C/Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 24-May-2001
 C/Accession: A90267
 R/She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Aways, M.J.; Chan-
 Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, F
 arett, R.A.; Ragan, M.A.; Seneen, C.W.; Van der Oost, J.

submitted to Genbank, April 2001
 A/Description: Sulfolobus solfataricus complete genome.
 A/Reference number: A99139

A/Accession: A90267
 A/Status: preliminary
 A/Molecule type: DNA
 A/Residues: 1-1269 <KUR>
 A/Cross-references: GB:AB006641; NID:g13814326; PIDN:AAK1392.1; GSPDB:GN00155
 C/Genetics:
 A/Genes: SSO1141

Query Match 3.7%; Score 125.5; DB 2; Length 1269;
 Best Local Similarity 19.6%; Pred. No. 2.4;
 Matches 98; Conservative 69; Mismatches 181; Indels 153; Gaps 27;

QY 18 RNRVRSANTVTYVNGRNORRRGRQVSPPDNTAAADLAQSLDANTVTPPANISMPER 77
 Db 584 QRLVTYANITYPPG-----SPVO--TGBRKALILEVLTGNTLTFTNLTYSNLTYL 629
 QY 78 RHWAKKIDLDSDSIGVFKYLDPAQATASARAVGEYSKIPDGVKESVDAREIY--- 134
 Db 630 --WT--GSGVLSNKAASILVTV--YSSDGLAGIGYETFPSTGYITFTTTPYVEL 664
 QY 135 -NEECVVTDS---VPLDGRQMSLISFSPMERETAYVAVAN---VENKENSIDVVDL 186
 Db 685 GNAELGITLSNSYFOAPIGVMTLTNTISYNITNATVTTLSVPIKNGVVIDLPDL 744
 QY 187 IEWNLNLAWRVYVDSQW--INFVDT--TYVRIYV-----LNP 223
 Db 745 -----SIGDLILAEKNAVGFDAFTNGVMQTLFILPQVVERGSVSPQHITIEGSIIP 799
 QY 224 TYDVPDPT--EGLVRTVSQYRLTYKATC-----RANMPTLVDOGFWIGQVALPTSL 275
 Db 800 PVNLPESTTQDALQGT---NITAKLVSSNGVYINENLP--LSPNIIYFG--VLYIPKNT 852
 QY 276 PQ--YDV-----SEAYVALHTLTPARPSAAL-----AFVWAGLPQ 309
 Db 853 PSLGVNVLLFATYVSYTLNTTIRGFYQGIYVSNQATISVKSVMNPAFEGQTVIVANTIN 912
 QY 310 CGTA-----PAG-----TPAEQASSGGYLTMRHNGT--TPPAGS 342
 Db 913 GTNIEIKGFMSATVYSSLSFNTYTTISSIIRIPLWNPKIG--EWEKNFTLPSAISAGN 969
 QY 343 VSVVLEPGALERYDNDQSWTDPASAGDTVFRQVAVBVVTTNPPAGGSAPTTVAV 402
 Db 970 LTYLACGGYF-----GVPEK-VLTIGISALGNPTTNSGNATYIN 1009
 QY 403 PPSNATNTVFRNLTLETPPS 423
 Db 1010 LPYTLFTN-----QTLDKTLP 1026

RESULT 6

T28698
 hypothetical protein - Streptomyces coelicolor
 C/Species: Streptomyces coelicolor
 C/Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 R/Parhill, J.; Bentley, S.D.; Barrell, B.G.; Ralmandream, M.A.
 submitted to the EMBL Data Library, January 1999
 A/Reference number: Z20512
 A/Accession: T28698
 A/Status: preliminary; translated from GB/EMBL/DBJ
 A/Molecule type: DNA
 A/Residues: 1-2314 <PAR>
 A/Cross-references: EMBL:AL023496; NID:e1292348; PID:e1292365; PIDN:CAA18915.1

Query Match 3.7%; Score 125; DB 2; Length 2314;
 Best Local Similarity 21.0%; Pred. No. 6.5; 252; Indels 260; Gaps 36;
 Matches 157; Conservative 77; Mismatches 157; Indels 260; Gaps 36;
 QY 35 ORRRRT--GRVSPDNFTAAADLAQSLDANTVTPPANISMPERFWAKGKIDLSDSI 92

Db 979 ERAATVLDASAPRTCTLMAAHDLASVADALTTDEGGVAVLQADPPDRMEGP-SLVKEA- 1036

QY 93 GWTFKYLDPAAGATESAPAVGEYSKI PDGLV-----KFSVDAREIRY 134

Db 1037 --YSSDDDLGGVTEGAWME--QYTDLDGFGVAPRSVSVRYANSGAAAPSSVDIHAD-- 1090

QY 135 NEBCPVVTDVSV-----LDGQWMSLSTF-- 156

Db 1091 -ADGPVAVATSLPCTGCMQYTTTVAASVDFQALLKASSATFVPHADSGROM-VSNPDWY 1148

QY 159 -----SPPMERTAVVAVANVENK--EMSLDVNDLLEMLNN--LADMRVYVSEQ 204

Db 1149 QFSPEAPSSSPITTLATLTANNTSTGDSGLPLKVGGEVENTNGAMEMR----- 1201

QY 205 WINTNDTTYVYVIRVLRPTVDPDPTEGLVTVSDYRLTK-----AICE----- 251

Db 1202 -----CTDLGDGADTAVSYDKPRS-----RAASDSHIELRPGAKDGPATVPLDYTG 1250

QY 252 -----ANMPTLVVQGFNIGQ--YALTPTSLEQYDVSEANA-LHTLTFRAPSSAA----- 298

Db 1251 SGWGTAVASTVRLDPDVEEGTQDYAVFVSS--TQTDAPYVAVNHSILTLTRQADAPVVPD 1309

QY 299 ALAFVWAGLPQGTAPAGTAPAMEQASGGYL-----TWRHNGTFFPAGSVSYLPEGFAL 353

Db 1310 ATAF-----EGSSGGGLKSEPAITSGAGSATSLGCT----- 1340

QY 354 ERYDPDGSWTDPAAGDTTFRQVAVDEVVT--NNPA--GGGSAPTFTVRVPSNAYT 409

Db 1341 -----YCGAMLDY---GD-YDFGDSPKKVTTLTYVNNASARCQTSAAV----- 1378

QY 410 NTVFRNTLTETRPSSRLLEPMPPADPGQTVANNPKIEQSLKELTGCYLVHSGMRNVF 469

Db 1379 -QLVLDSPDPAPGTPYATVPLPVT--GSSWSGGTSTLTPLEAITGHAHLELTTNAD 1435

QY 470 QLTP--ASSFGAVSFNNPGYERTDLPDYTGIRDSFDQMSFAVHFRSLSHSCSIVTKY 528

Db 1436 SSHYVYVNLGOVALDRVAPAKOTDL-----SALKRAI 1467

QY 529 QGMEGVTN-----VNTPEQGF-----AHAGLLKNEELCLADLATR-LTGVYPATDNFA 577

Db 1468 EQYGLSEADADRYGTIDGVRFRRLTARLDLTGEDATQLBADLTRSLT----- 1517

QY 578 AAASFAPANML-----SSVLKSEATSSIIKSGETAVGAAGGLAKLRPL 623

Db 1518 -----LAAHQVLPRLRLLESIVATASALADERYTDASWKAFTTALTAAKTALADERTATD 1572

QY 624 MSVPKLIARVARARA--RRRAARA 646

Db 1573 RTLTERVALLDRARSSLTTKKRTVPA 1598

RESULT 7

C75015

Probable: pyrolysin (EC 3.4.-.-) homolog PAB1252 [similarity] - Pyrococcus abyssi (strain C:Species: Pyrococcus abyssi

C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000

C:Accession: C75015

R:anonymou, Genoscope

submitted to the EMBL Data Library, July 1999

A:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: C75015

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-1204 <KAM>

A:Cross-references: GB:A0248288; GB:AL096836; NID:g5458960; PIDN:CAB50561.1; PID:g545907

A:Experimental source: strain Ozeay

C:Genetics:

A:Gene: PAB1252

C:Keywords: hydrolase

Query Match 3.7%; Score 123.5; DB 2; Length 1204;

Best Local Similarity 17.4%; Pred. No. 3;

Matches 145; Conservative 105; Mismatches 261; Indels 323; Gaps 45;

QY 46 PDNFTAAQDLAOSLDANTVTPPANISMPERFNNAKGIDLDSDSI-----GWTRYLTD 100

Db 237 PENYAVWVWPLTQWMTA-----PHN--SGNPBMTFWO--LYDSPDSLYNNPGDWYILYLD 288

QY 101 -----PAQTSARAV--GEYSKIPDGLVPSVDAIR--EIVNEBCPVV 142

Db 289 NETQWNYLKBLLKYGVYIQTAIIYVPSFDLHNPEDIVG-RLDKRYRLSKNSIDMTWTTG 347

QY 143 DVAVPLDGRQMSLSIFPMPFRATVAVANVENKESLDVNDLI----- 187

Db 348 DVA-----DMTRVL--LSWVGTAYGNINGRITISIELMCKVLTFRDPCVSLDPTIPEA 400

QY 188 -----EWLNLTADR-----YVDSQWINFND--TTYVRIIRVLRPTVYD 227

Db 401 VQKINATPYTRFVNNETWEKASPLYSTSLRGEW--FVNNAFVDIYALLNFMMLYYI 458

QY 228 PDPF--EGLVRVSDYRLTYKAITGBANPILV----- 258

Db 459 PEGTYVKSGRINYI-DPLTFYQMT--GHAVTIIGYDDNTTTPDGKALIMVNSKGTWNG 516

QY 259 DOGFWIGQYALTPTSLEQYDVSEAYALHTLTPARP-----SSAALAFV 303

Db 517 DNGFW--KYS-----YKARTTYHDYRGNPYVISVRDIISLPVFIISVGEAFVYV 565

QY 304 -----WAGLPQGTAPAGTAPMEQ-- 322

Db 566 PKAADYPRKLTMTVVGIGHPRGHVIDGVNATNYBILYMAGIIV--GVAVNGSPVMEHSFL 624

QY 323 -----ASGGYLTWRHNGTTPAG 341

Db 625 DFMMDYISDBELSIKQNVNEISILPQAHPPDPSMAPRVSEADITIMEILSESNTPL 684

QY 342 SVS--YVLPREGFALERY-----DPN--DGSWT----- 364

Db 685 TVDPYVAPKQILADNFGVLYNFTVLNINGBLKVLGSLDKNYSIPGSGWYIPLLEIPV 744

QY 365 DPASAGDTVFR--OAVDEVVVTNNP--AGGSAPT-----FTVRVPSNAY 408

Db 745 SYDAFANNVITIKYGDFFNVSYISIIPLGAKVYIGKSEYPKAKBEGGYYYGTBIAEKLEL 804

QY 409 TMTVFRNTLTETRPSSRLLEP-----MPADPGQTVANNPKIE--OSLKET 454

Db 805 SAGTYNTVAVVTPNGKXVALPSRIYVITISGPTIKIYSPVD--TVNVVTTIPIELVNH 861

QY 455 LGCTLVHRSKRNIPVQLTTPASSFGAVSPNNPGY--ERTDLPDYTGIRDSFDQ--NMST 509

Db 862 LKINVTAAVGGKSIELT-----YNAIS--GYTGKVTILENGAYTLVVTATBELNANVGT 913

QY 510 AVAHFRSLSHSCSIVTKYQGMBCVTNVNTPFGQPAHAGLLKNEELCLADLATRLT-- 567

Db 914 ARVHR--VVSANAKVTPVTEN--TYVT-----VGVTGNATITVVTDTVAVANTTS 960

QY 568 -GVY-----PATDNFAAVSAFA--ANMLSVLKSEATSSIIKSGETAVGAAGS 615

Db 961 EGTEYKVPVYNN--APALFNSTALBDDVAKGAAHATLVAGMNVVTTTKTBVG 1011

RESULT 8

S29605

Glycoprotein 350/220 - human herpesvirus 4

C:Species: human herpesvirus 4, Epstein-Barr virus

C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 26-Aug-1999

C:Accession: S29605

R:Klein, K.; Mueller-Lantzech, N.

submitted to the EMBL Data Library, October 1992

A:Description: Sequences of the membrane proteins gp 350/220 and p140 of Epstein-Barr

A:Reference number: S29605

A:Accession: S29605

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-886 <KLB>

A:Cross-references: EMBL:X67776; MID:959163; PIDN:CAA47986.1; PID:959164
 C:Superfamily: Epstein-Barr virus membrane antigen gp350
 C:Keywords: glycoprotein

Query Match 3.6%; Score 123; DB 2; Length 886;
 Best Local Similarity 20.3%; Pred. No. 2;
 Matches 136; Conservative 66; Mismatches 269; Indels 206; Gaps 29;

```

QY 48 NPTAAADLQSLDANTVTPPANNISSMEPRNMAKGIKDLSDSICGTFKTLDPAGATES 107
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 169 NITAVVR--AQGLD---VTLPLSLPTSADSNFS---VTEMLG---NRID---IBC 211
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 108 ARAVEYSKIPDGKVFESVDAIRE-----INNECPVTVDVSVPLDGRMSLSIFSF 160
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 212 IMEGEISQVLPGRNKNITCSGYESHVPSGGLILTSIPVAT--PIRGTAYLSRLRPR 269
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 161 PMFR-----TAYVAVANVENKMSLD-----VNDLI---EWLNILADMYVVDSE 203
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 270 PVSFPLANNSTILYFVSGNGPKAGGDYCIQSNIVFSDIIPASQDMPTNTDITVVDN- 328
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 204 QMINPTDITYVAIRLRLPTVDYDPDEGLVRLVSDYRLTYKAITCEANPTLVDOGFW 263
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 329 -----ATYVSPMVTSE-----DANSPNTVTAFW 352
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 264 -----IGQYALPT----- 273
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 353 AMPNNTETDFKCKTLTSGTPSGCENISGAFASKRTPDITVSGLTAKTLITRTATNA 412
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 274 -----SLQYDVSEAVLHTLTPARPSSAAALAFWMAGLPQGGTAPAGTPAMQAS 324
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 413 TTTTHKVI FSKAPBESTTSPPLNTGTFAAPMT-----TGLPSTHPTNLTA--PAS 463
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 325 SCGLVTRHNGTTPASVSIVLPEGFALERYDPNDG-----SWDPASAGTPTVTRQ 377
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 464 TGPVTSADVTSPTPAGTSGASP--VTSPSPRDNGTESKADMTSPSAVTPPENNA 520
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 378 VADEVVVT-----NNPAGGSAPTFTVRVPPSNAYNT-----VFANLTLETSPS 423
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 521 TSPPTATTPTPNATSPITLGTSTSAVTTTPNATSTPAVTTTPNATITPTLGTKISPT 580
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 424 SRRLPMPADFCQTVANNPKIEQSLKETLGCYLVHSMKRNVPFOLTPASFGAVSFN 483
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 581 S-AVTTTPNATSPVGETSP--QANTTNHTLG-----GTSTPVVTSPPNATSAVTTG 632
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 484 NPGKERDLPDYTGIR-DSFDQMSSTAVAHFRSLSHCSIVTQYQMEGVTVNTPFG 542
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 633 Q--HNITSSSTSMSLRPSISETLSPSTSD--NSTSHMPLLTSAHPTGEMITGV--TPAS 688
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 543 QPAHAGLLKNEEILCLADLALTRLTGVPAT-----DNFAAVSAPFAANMLSSVLKS 595
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 689 TSTH-----HSTSSPAPRPAGTTSQASGPNSSSTTKGGEVNVTKGTPPKN 734
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 596 ATSSIIKSVGETAVGAQS 614
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 735 ATSPQAPSGQKTAIVPTVS 753
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 9

AC1153
 addressin homolog lmo0627 (imported) - Listeria monocytogenes (strain EGD-e)
 C:Species: Listeria monocytogenes
 C>Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AC1153
 R:Glaser, P.; Frangoul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, D.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgat, O.; Entian, K.D.; Fahl, H.; Jones, L.M.; Karet, U.
 Science 294, 849-852, 2001
 A:Authors: Krefit, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maicouram, A.; Mok, C.; Schlueter, T.; Simoes, N.; Tiller, A.; Vazquez-Boland, J.A.; Voos, H.; Weiland, A.; Title: Comparative genomes of Listeria species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AC1153
 A>Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-1582 <GTA>
 A:Cross-references: GB:NC_003210; PIDN:CAC98705.1; PID:g16410016; GSPDB:GN00177
 A:Experimental source: strain EGD-e
 C:Genetic8:
 A:Gene: lmo0627

Query Match 3.6%; Score 122; DB 2; Length 1582;
 Best Local Similarity 19.6%; Pred. No. 5.9;
 Matches 122; Conservative 92; Mismatches 247; Indels 160; Gaps 35;

```

QY 31 GRNRRTGROVSPPNFTAAADLQSLDANTVTPPANNISSMEPRNMAKGIKDLSD 90
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 31 GAENQKTKQQAQAPDQKADBN-----KTTVPDKERT--KLVNKG-QLLKG 79
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 91 SIGMYF--KYLDPAGATESARAVEYSKIPDGKVFESV-----DARIREIYN----- 135
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 80 SYDMNFTNKVTDANGTKETAPCD-----SLKFSVSLAIPESBAVLDETFTFWIQK 132
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 136 --EBCPVTVDSVPLDGRMSLSIFSPMERTA-----YV-----AVAN 172
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 133 DPFQGDVTVNPAIDELTPRQNVFTDMSKNSIGTGTINGVEYIGYTLKFNOSPEALAK 192
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 173 VENKENS---LDVNDLIEWLNILADMR--YVDSQWINFNTPTTYVRLRLPTVDVP 228
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 193 YEDEPLNACPDMEPATISTITTKEDYKAIVIEDHVDISN-----IEVPVTPPT--P 244
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 229 DPTGLVRLVSDYRLTYKAITCEANPTLVDOGFWIG--GQYALPTSLPQYDVSEAYA 285
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 245 DTONG-VLTAS-----KVIDSAMRLDKDQGVKVKVATSTNTYPEKODL 287
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 286 LHTLTPAPSSA--AALAIPWAGLPQ--GTRAPGTAPMEQAS--SGGYLR--W--RANGT 336
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 288 VPTVYVAMNNSGENAVIERKIDELPAFSLITTESPFWQALSTASGSLRGWVASADQT 347
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 337 TTPAGSVSYVLPBGFALERYDPNDGSWTDPASAGDTVTPQVAV-----DEVVYNNP 389
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 348 NMAPGARLYLPPSPA-QTISKNKG-----YTTIAKVTDPKDLTNVTKSP 395
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 390 AGGGSAPFTVRVPPSNAYNTVFRNTLLETSPSRLBLPMPADFCQTVANNPKIEQS 449
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 396 TKGEGKSGISGPSKQVY-----DWAITTKISS-----TYDKNNKX--- 432
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 450 LKETLGCYLVHSMKRNVPQULTPASFGA--VSFNNPGYERDLPDYTGIRDSFQNM 507
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 433 -----IGAV--SSTTTPV-SISEGSTVGTYTAINGNY--TKDVVVVAVVPAGYEILAN 481
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 508 STAVAHFRSLSHCSIVTQYQ--WEGVTNNTPPGQPAHAGLLKNEEILCLADLAT 564
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 482 DT--SDFKTANDKKKTEITVPSGGVMSDIAKVTSITLSCMASGATKTAATMKA--MGI 537
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 565 RLTVGPATDNF-AAAVSAPA 584
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 538 PEDQTVSAVDPFYMAGEIGSPS 558
   |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||

```

RESULT 10

Q08B21
 membrane antigen gp350 - human herpesvirus 4 (strain B95-8)
 C:Species: human herpesvirus 4, Epstein-Barr virus
 C>Date: 25-Feb-1985 #sequence_revision 25-Feb-1985 #text_change 28-Jul-2000
 C:Accession: A43042; G33008; G33009; A03762
 R:Bankier, A.T.; Deininger, P.U.; Farrell, P.J.; Barrrell, B.G.
 Mol. Biol. Med. 1, 21-45, 1983
 A:Title: Sequence analysis of the 17,166 bp EcoRI fragment C of B95-8 Epstein-Barr virus
 A:Reference number: A93065; MUID:55035713; PMID:6092825
 A:Accession: A43042
 A:Molecule type: DNA
 A:Residues: 1-907 <BAN>
 A:Cross-references: GB:V01555; GB:J02070; GB:K01729; GB:K01730; GB:V01554; GB:X00498; G
 R:Farrell, P.J.
 submitted to the EMBL Data Library, March 1986
 A:Reference number: S32973

A:Accession: S33008
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-907 <FAR>
 A:Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24855.1; PID:g1334869
 A:Accession: S33009
 A:Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-907 <FAR>
 A:Cross-references: EMBL:V01555; NID:g59074; PIDN:CAA24855.1; PID:g1334869
 R:Baer, R.; Bankier, A.T.; Biggin, M.D.; Deininger, P.L.; Farrell, P.J.; Gibson, T.J.; H
 Nature 310, 207-211, 1984
 A:Title: DNA sequence and expression of the B95-8 Epstein-Barr virus genome.
 A:Reference number: A03794; MUID:84270667; PMID:6087149
 A:Contents: annotation; protein coding region
 C:Superfamily: Epstein-Barr virus membrane antigen gp350

Query Match 3.6%; Score 121.5; DB 1; Length 907;
 Best Local Similarity 19.6%; Pred. No. 2.7; Mismatches 266; Indels 233; Gaps 28;

Matches 138; Conservative 66; Mismatches 266; Indels 233; Gaps 28;
 48 NFTAADLAQSLDANTVTFPANISSMBEPRMAGKIDLDSDSIGWFKYLDPAQATES 107
 169 NITAVR--AQLD--VTLPLSLPTSAQDSNFS-----VKTEMG---NEID-----IEC 211
 108 ARAGEYSKIPDGLVKSVDARE--YNEBCPVYTVDSVPLDGRWMSISFSF 160
 212 IMEDDEISQVLPFGDNKFNITCSGYESHVPSCGILSTSPVAT--PIPGYAVSIRLTPR 269
 161 PMFR--TAYVAVAVENKEMSLD-----VYNLDI--EMLNLDLMDRVVVDSE 203
 270 PVSRLGNNSILIVYSGNGPFAAGDYCIQSNIVSDEIPASQDMPTNTDITVGDN- 328
 204 QMINFTNTYTVRIRVLRPTVDVDPTEGLVTVSDVRLTYKAITCEANMPTLVDOGF 263
 329 -----ATYSVMWTSE-----DANSNVVTVTFM 352
 264 -----IGQVALTPT----- 273
 353 AMPNNTETDFKCKWTLTGSPSCENISGAFASNRTFDITVSGLTAPKTLITRTATNA 412
 274 -----SLPQYVSEAYALHTLTPARSSAALAFVWAGLPQGTAPACTPAMEOAS 324
 413 TTTTHKVIIFSKAPBSTTTSPTLNTGTGADPNTT-----TGLPSTHVPNLTN--PAS 463
 325 SGGVLTWRHNGTTFPAGSVSYLPEGFALERYDPNDGSV-----TDFASAGPTV 373
 464 TGPTVSTADVSTPTAGTTSASP-----VTPSPFMDNGBESKAPDMTISSTSPVTP 516
 374 TFRQVAVDENVVT-----NNPAGGSAFTFVRVPPSNAVTVYFR 414
 517 TPNATSPPTAVTTPPNATSPPTAVTTPPNATSPTLGTSPTSAVTTPPNATSPTLGK 576
 415 N-----TLLETPSSRRLLEPMPADRCQTAAANPKIEQSLKXETLQCY 458
 577 TSPISAVTTPPNATSPTLGTSPTS-AVTTPTPAT--GPTVGETSP--QANVTNHTLG-- 631
 459 LVHSKRRNPVQLTPASSFGAVSFNNPGYERTDIPDYTGIRDSFDQNMSTVAHFRSIS 518
 632 ---GTSPTPVVTSQPKMATSATTTGQ--HNITSSSTSSMSLARPSNPETLSPSTSDN 686
 519 HSCSIVTKTYQGEVGVNVTNPFQGFANAGLKNBEILCLADDLATRLTGVPAT----- 573
 687 HNPILTSAPHTGGENITGV--TPASISTH-----HVSSTSPAPRPPTTSQAS 731
 574 ---DNFAAVSFAFAMLSVLKSEATSSIIKSGETAVGAAS 614
 732 GFGNSTSTKGEVNVTKTTPONATSPQASGQKTAFTVTS 774

RESULT 11
 B75514
 penicillin-binding protein 1 - Deinococcus radiodurans (strain R1)

C:Species: Deinococcus radiodurans

C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000

A:Accession: B75514
 R:White, O.; Bisen, J.A.; Heidelberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
 M.; Shen, M.O.; Venter, J.C.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;
 S.; Smith, H.O.; Fraser, J.C.; Fraser, C.M.
 Science 286, 1571-1577, 1999

A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
 A:Reference number: A75250; MUID:20036896; PMID:10567266

A:Accession: B75514

A:Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-873 <WHI>

A:Cross-references: GB:AE001907; GB:AE000513; NID:g6458162; PIDN:AAFI0059.1; PID:g6458;
 A:Experimental source: strain R1

C:Genetics:
 A:Gene: DR0479

A:Map position: 1

Query Match 3.6%; Score 120; DB 2; Length 873;
 Best Local Similarity 26.0%; Pred. No. 3.3;

Matches 72; Conservative 19; Mismatches 104; Indels 82; Gaps 16;

208 FTNDTTYVVRIRVR-----PTYDVPDP-----EGIVRVSDVRL 243
 545 FANGGLIYAPSAVRMBDPQGRVLYKRPDPVGRVMDKRAMGLDMIRGVNDLSAYQ- 603
 244 TYKAITCEANMPTLVDOGFWIGQYALTPTSLPOYDVSAYVALHTLTPA--RPSAALA 301
 604 --GGATPARR-----DGMVQGGKTGTT-----NDVVD-----LWFAGTPTLSGA- 642
 302 FYVAGLPQGGTAPR-----GTPAMEOASGGVLTWRHNGTTPAGSV-----SVLPBG 350
 643 -VWVGKQGGTLPSPMAVSGVPTPVW--QAAAGLAGLRETATPTPGGIVVRYRYRLMA 700
 351 PALRBYDP-----NDGSMWTFASAGDTVTFROVADE--VVYTNPNPAGGSAP-----T 397
 701 FLTERADDPVRHDS-----RRAITYKATTPAVTVAPBPQAPVTTTPAGSPAPLRRRTRT 756
 398 FTVRVPPSNAYNTVFRNTLTLETPSSRRLLEPMPRA 434
 757 LPRGTSRPRALPS-----RTGPRRPLSPPPRRPRGA 769

RESULT 12

T09055

protocadherin 68 - human
 C:Species: Homo sapiens (man)
 C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 11-Jun-1999

A:Accession: T09055
 R:Jin, P.; Xu, H.; Israel, D.
 submitted to the EMBL Data Library, October 1997

A:Reference number: Z16540

A:Accession: T09055

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-889 <JIN>

A:Cross-references: EMBL:AF029343; NID:g2599501; PID:g2599502
 C:Genetics:

A:Gene: PCH68

Query Match 3.5%; Score 119.5; DB 2; Length 889;
 Best Local Similarity 20.4%; Pred. No. 3.6;
 Matches 144; Conservative 92; Mismatches 250; Indels 219; Gaps 36;

32 RRNORRRRGROVSPDNTAAODLAQSLDANTVTFPANISSMBEPRMAGK----- 83
 97 RHNAKQSLSLVFPANDKEICIKIVRIDINNAFSFSSDQXLDISNAAGCTPPLTSA 156
 84 -KIDLDSDSIGWPKYLDPAQATR--SARAVGEYSKIPDGLVKSVDAREIRIYNEBCPV 141
 157 HDPDAGENGARTYLLTRDHLFGLDVKSRSDDGRTKFPFLVIQKLD---RQGNHHTLV 213

```

QY 142 TDVSVPLDGRQMSLSIFSEPMFRTAYAVANVENKENS--LDVNDLIEMIANLADRRYV 199
DB 214 T-----ALDGB-----PPRGATVOINNAVINDNSNPVEAPSLVEIPENTPIGTV 262
QY 200 VDSEQMINFTN-----DTTY-----YVRIRVLRPTVDPDPTEGLVTVS--DY----- 241
DB 263 ID-----LNATDADEGNNGEVLVGFSSVYVDRV--RELPSI--DPKGTGLIRVGNLDYENGM 316
QY 242 -----RITYKAITCEANMPTLVDOGFWIGGQVALTPTSLPQVDVSE 282
DB 317 LEIDVQARDLGNLIPAHCKVTVKLIIDRNDNAPSI--GF-----VSVRQGLSE 363
QY 283 AVAALHTLTFAPSSAALAFVW--AG-----LPQGTAPAGTAMEQSSGGLT 350
DB 364 A-----APGETVIALVRVTDNRDSDGKNGQLCCRYLGGGIGGAG----- 401
QY 331 WRHNGTTPPAGSVSYVLPESFALERYPDNDGWTDFASAGDTVTFQV--AVDEVVTN 387
DB 402 -----GLGPGGGSVPFLLENY-----DNFYT-----VYTDHPMDEFTQDEYVNTI 442
QY 388 NPAGGGSAP-----TTTVRV-----PP--SNAVYTVFRNTLLETRESRLLELMP 433
DB 443 VARDDGSPPLNSTKSPAIKILDXNDNPPRETKGLVYLQVHENNI-----PG 488
QY 434 ADEGQTVANNPKIEO-----SLKKTIGCYLVHSMKR--NPVQLTPASSFGAVSRNPG 486
DB 489 DKLGSVLADQPDGQNGTVSYSLPISHIGVSYITYVSVNP-----TNGALVALRSFN-- 541
QY 487 YERTRLDPYTGIRDS-----PDQNMSTAVAHFRSLSHSCSIYTKTYQWEG----- 533
DB 542 FEQTKAFEFKVLAKDGGAPAHLESNATVRYTVLDVNDNAVILPILQNTALQVPRNA 601
QY 534 -----VTNNTPPGQFAHAGLK-----NEBILCLADLIALRLTGVPATDN-----F 576
DB 602 GLGYLSTVRALDSDGESGRLTYEIVDGNDDHLFEIDPSSGEIRTLHPWEMEDVTPVEL 661
QY 577 AAASVAFANMLSSVLKSEATSSIIKSVGETAVGAQAGSLAKIPG 621
DB 662 VVKVTHGKPTLSAAVK-----LIRSVS-----GSLPEGVPRNG 697

RESULT 13
P90073
hypothetical protein SA2447 [imported] - Staphylococcus aureus (strain N315)
C/Species: Staphylococcus aureus
C/Date: 10-May-2001 #sequence_revision 10-May-2001 #text_change 22-Oct-2001
C/Accession: P90073
R.Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc,
me, A.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.
Lancet 357, 1225-1240, 2001
A/Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.
A/Reference number: A89758; WUID:21311952; PMID:11418146
A/Accession: P90073
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-2271 <KUR>
A/Cross-references: GB:BA000018; PID:G13702612; PIDN:BA843752.1; GSPDB:GN00149
A/Experimental source: strain N315
C/Genetic:
A/Gene: SA2447

Query Match 3.5%; Score 119.5; DB 2; Length 2271;
Best Local Similarity 18.9%; Pred. No. 15;
Matches 147; Conservative 107; Mismatches 233; Indels 289; Gaps 37;

QY 22 VSANTYTVN-----GRNQRRRTGRQVSPDNFRAAODLADLSDANTVTPPAN 70
DB 254 VTANITTVKNDLKQYMTTSGNATYDOSTIGITVLTDAYS-----QKGAITVLGTR 303
QY 71 ISSMPEFRWAKGKIDDS-----DSIGWYFK----- 97
DB 304 IDSNKSFH--FSGKVLGNKRYEGHNGGDOIGFAFSGVLTGETLNGAANGIGLSNAG 361

```

```

QY 98 -----YLDPAQATESARAVGYSKI-----PGGLYKFSY 126
DB 362 FKIDYHNTSKSPKSAKANAADPSNAGGAPAFVTTDSYGAVTYTSSTANMAKLV 421
QY 127 -----DAIRIENESCPVTVDSVPLDGRQMSLSI-----FSPPM----- 162
DB 422 QPTNNTFQDPDIN--YNGDTKWT--YKAGQTVTRNLSDMYAKSGTTNFSLSMTRASTG 476
QY 163 -----FRT--AYVAVANVENKENSLDVNDLI--EMIANLADRRYVVDSEQ----- 204
DB 477 GATNLQOVQFGFTREYTESAVTVQVRYDVTGKDIIPKTYSGNVQDVQVITDNGQSLATK 536
QY 205 WNFPTN--DTTYVRIARIVAPTHDVPDPTEGLVTVSDIRLYKAITCEANMPTLVDOGF 263
DB 537 GNYVTSVDSY-----ASTVN--DTNKTVKTNAGQSTVFTYTVKAKVTVVAGQOTIE 587
QY 264 IGGQVALTPPLSLPQVDSBAVALHTLTPARPSAALAFVWAGLPQG-----GTAAGT 317
DB 588 VG--KTMNPFVLLTTNGTGTVNTIVT-----GLPGLSTYDSATNSIIGT 630
QY 318 P-----AMEQASGGYLTMERN--GTFPAGSVSYVLPBG--PALERYDP----- 358
DB 631 PTKIGQSTVTVSTQDANKKSTTFITINVDTTAF-----VPIDQSESVYSPIPIK 685
QY 359 -----NDGWTDPASAGDTVTRQVAVDEVYVTTNPAAGGSAFTF-----TRVPSNAY 408
DB 686 IATQDMSGN-----AVNTVYTGLPSPGLTFPSTNNTTIGTPTNIG 724
QY 409 TMT-----VFRNTLLETRESRLLELMPADPQGTVANNPKIEQ 448
DB 725 TSTISIVSTDASGNTKTTTFFKTEVTRNSMSDSVSTGSGTQ-----SGSVSTGADSQ 777
QY 449 SLKKTIGCYLVHSMKRNPVQLTPASSFGAVSFN--NPYERTRLDPYTGIRDSFDQNM 507
DB 778 SASTSGSGLVGS-----TSASTGKSSTSVLSDSVASAKSLIS--TSBENSVSST 825
QY 508 STVAHFRSLSHSCS--IVTKTYQWEGTVNTNTPQOPAHAGLKNKEE1----- 555
DB 826 STLVNSQGVSSMSGSVSKSTLSDSISNSN-----TEKESLSSTSDSLRTS 876
QY 556 LCLADLIALRLTGVPATDNFPAAVSAPANMLSSVLKSEATSSIIK--SVETA 608
DB 877 TSLSDLSMSTSGSLKSGSLSTISGSSS--TSALSDSTSNALSTSLSESA 929

RESULT 14
T43220
Insulin-like growth factor-1 receptor - common lancelet
N/Alternate names: Insulin-like peptide receptor
C/Species: Branchiostoma lanceolatum (common lancelet)
C/Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-May-2000
C/Accession: T43220
R.Pashmforough, M.; Chan, S.U.; Steiner, D.F.
Mol. Endocrinol. 10, 857-866, 1996
A/Title: Structure and expression of the insulin-like peptide receptor from amphioxus.
A/Reference number: Z23446; WUID:96408719; PMID:8813726
A/Accession: T43220
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-1163 <PAS>
A/Cross-references: EMBL:883394; NID:G1911771; PID:G1911772; PIDN:AA850846.1
C/Superfamily: Insulin receptor; protein kinase homology
C/Keywords: hormone receptor

Query Match 3.5%; Score 119; DB 2; Length 1363;
Best Local Similarity 19.8%; Pred. No. 7.6; Indels 180; Gaps 23;
Matches 99; Conservative 60; Mismatches 162;

QY 79 NNAKGKIDDSDSIGWYFK-----YLDPAQATESARAVGYSKIPLDGLVKSVD 128
DB 416 DMDRTDITIDBGLTFPHFNPKLGRHYILTMVDKVGLEPRA-----ITD----- 458

```

QY 129 EIREIYNECPVTVDSVPLDGRQMSLSIFSPMERTAVVAVANKEMSLDVNDIE 188
 DB 459 -----TDISTLTNGDQ-----AQCSPSRLE-----IE 480
 QY 189 WLNLLADRWYVDSBOWINFTNDTTYVYIRIVLRPTVDPPTBGLVTVGDRLYKAI 248
 DB 481 EINTSKMIL-----RMSF-----RP-----PDRDLSTTVS-YRET----- 514
 QY 249 TCEANMPTLVDOGF-WIGGOYALTPTSLPDYVSEAVYALHTLTPARPSAAL--AFVW 304
 DB 515 -----EDQGIDEYGDACGNTBKEFDVSPQIAHITGLKPTQYALLKTYTK 565
 QY 305 AGLPQGS-----TAPACTPAMEQSSGCIYLTWRNNGTTPFAGSVSYLP 348
 DB 566 AGAREGSGAKSDIYVARTADAKPTHPQDVVYVSNSSNTLIITWKPNR--PQNGVTHYIV 623
 QY 349 EGFALERYDPNDGSMTPDASAG-----DTVTFROVAVDEVYVYVNNPAGGSAFPTVAVP 403
 DB 624 K-YKQEDVAVMEQREYCKGLKPHRPTQGLIEDIVNNEEPNNSTIDGTC---CECP 678
 QY 404 PS-----NAYTVFRNTLLETRPSSRLLELPMPADP--GQTV--- 440
 DB 679 KSEDEIRIEEBEAAFOGEFENFLHNHYHKKENETRAGRRELPVLTARPFYSNQTAVVT 738
 QY 441 -----ANNKIEOSLKEITLGCYLVHKKRN--PVFQITPSSFGAVSFN 483
 DB 739 LPSTRTVPPTPTPNPNQLETTVNNHM--VVLTLGRHSEYILIEVIACNADAAVCS 795
 QY 484 NPGYERTDLPDYTGIRDSFD 504
 DB 796 GSAVELARTQAD-----DSAD 811

RESULT 15

AF1450
 Probable peptidoglycan bound protein (LpxNG motif) lin0141 [imported] - *Listeria innocua*
 C:Species: *Listeria innocua*
 C:Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 27-Nov-2001
 C:Accession: AF1450
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, J.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussauget, O.; Entian, K.D.; Fehl, H.; Jones, L.M.; Karet, U.
 Science 294, 849-852, 2001
 A:Authors: Krett, J.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Meok, C.; Schlueter, T.; Simoes, N.; Tietzer, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland, A.; Title: Comparative genomics of *Listeria* species.
 A:Reference number: AB1077; MUID:21537279; PMID:11679669
 A:Accession: AF1450
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1993 <GLA>
 A:Cross-references: GB:AL592022; PIDN:CAC95374.1; PID:g16412560; GSPDB:GN00178
 A:Experimental source: strain Clp11262
 C:Genetics:
 A:Gene: lin0141

Query Match 3.5%; Score 119; DB 2; Length 1993;
 Best Local Similarity 19.3%; Pred. No. 14;

Matches 152; Conservative 100; Mismatches 276; Indels 260; Gaps 38;

QY 3 DAGVASQPHNRKGRNVRVSANTYTVNGRNRQRRRTGRQVSPPNFTAAADLAQSLDA 62
 DB 1247 DASVA-----TAKTLTDSVAIVGGSTN-----GKIRVTD-LTIALDIDIGTEKK 1291
 QY 63 NTVTTPANISMPERFNAKGI-----DLSDSICGYFYLDPAQATESARAVGEYSKIP 118
 DB 1292 STA-----PLYVTANGKVAQDEIKIGATNTFLBYKDAVVLSE-AQAVTK----- 1336
 QY 119 DGLVFSVDABIRIYNECPVTVDSVPLDGRQ-----WSLSIFSPMERTAVVAVAN 172
 DB 1337 ---AKTAAFEEYKGNVNSDAEDRTDV-VKVDQQLKALKNGSNRGTYPL---FTLEKN 1389
 QY 173 VENKMSLD--VVNDLIEMNLAD-----WRVYVDSBOWINFTNDTTYVYIRIVLR 222

DB 1390 GKEVEYIVDKVERDQLE--VNAHDSITTYGDNNRPAADNPFASALNKEGETLTFADIENTG 1447
 QY 223 P-----TYDVPDPEGLVTVSDYRLTYKAITCEANMPTLVDOGFMI----- 264
 DB 1448 TVDTTKAGBYVYTKNDITKTIYIYKD-----DAIBINAHDSITTYGDNWSAKDNDS 1502
 QY 265 -----GGOYALTPTSLPDYVSEAVYALHTLTPARPSAALAFWAGLPQGTAPACTPA 319
 DB 1503 AADRGENEVALSKTYVNTVNTAQAQGYPIYVT-----YGVGSKTITVT 1546
 QY 320 WEOASSGCIYLTWRNNGTTPFAGSVSYLPBGFALERYDPNDGSMTP--DPASA-----GDT 372
 DB 1547 VKENKKG--INAHNATITYG-----DSWTAEDNPFNAVDKQGNP 1583
 QY 373 VTFROVAVDEVV-VTNNPAG-----GGSAPFTVAV--PPSNAYTIVFRNTLLE 419
 DB 1584 VEFKKTVYTRPNNTKAGTYQAKTTPDASKTIVTLVKNICQVAVNANSTVYVGETWE 1643
 QY 420 TR--PSSRLLELPMPADPQGTVA--NNPKIEOSLKEITLGCYLVHKK-----M 464
 DB 1644 AKDNFDSARNKD-----GETVAFADVEEGVDMTVAGTVSITYKYDGFSTIKYTV 1695
 QY 465 RNPVQULTPASF-----GAVSFNPPG----- 486
 DB 1696 KNPQTALTADSVVYTGDNMSAKDNFDSALDKAGKPAVARDITVEBDPTVLDNTPGYTSV 1755
 QY 487 ---YE-----RTRDLPDYTG-----IRDSFDQ-----NMSTAVAH 513
 DB 1756 TYTKOGISKVQITVXPRQTVESHDSTIYAGASWAKNDNDSALDKKGDVYKLSDVYVI 1815
 QY 514 FRSLSHSCSIVTKTYQMEGVTNNTPPQGFARHAGLLKNEEILCLADDLATRLTGVPAT 573
 DB 1816 GRVDNQTPTGYEITYR-YDGVTSVSRVTVLQNHAKIIVNSKL-----KINADWPAK 1866
 QY 574 DNFAAASAPFANMLSSYLKSEBATSIIKSGEYAV-----GAHQSLAKU--PGL 623
 DB 1867 DNFRANSRDGSBIKPSKVKVGEKVN-TKAGAKYQVYTTTIDPNEGTVADAGEKBLVTATI 1925
 QY 624 MSVPQKIA 631
 DB 1926 EYVEGKVA 1933

Search completed: January 15, 2004, 16:53:08
 Job time : 16.9345 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 13:22:26 ; Search time 8.80647 Seconds
(without alignments)
3454.993 Million cell updates/sec

Title: US-09-991-262-50
Perfect score: 3374
Sequence: 1 MGDAGVAGSQRPHNRGTRNV.....GKIAARVAPRRARRRRAAAN 647

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	3.9	708	1	VP40_HCMVA
2	126	3.7	5703	1	MUSB_HUMAN
3	123	3.6	886	1	VGP3_EBVA8
4	121.5	3.5	907	1	VGP3_EBV
5	119	3.5	1363	1	ILPR_BRALA
6	115.5	3.4	699	1	CH11_BACCI
7	115	3.4	1419	1	ALAI1_CANAL
8	113.5	3.4	781	1	COAT_PAVHB
9	113	3.3	969	1	SACB_STRSL
10	112	3.3	1090	1	GUXB_CELFI
11	111	3.3	1902	1	P1P_IACLC
12	110.5	3.3	505	1	GUUV_ERWCA
13	109	3.2	591	1	FLGE_CAUCR
14	108.5	3.2	444	1	GUIN_ERWCA
15	108	3.2	775	1	VP4_ROT3
16	108	3.2	1260	1	ALSI_CANAL
17	107.5	3.2	2090	1	N214_HUMAN
18	106.5	3.2	609	1	HSE_SCHPO
19	106.5	3.2	776	1	VP4_ROT6
20	106.5	3.2	2201	1	POLG_CXA9
21	106	3.1	1138	1	TIF1_HUMAN
22	105.5	3.1	390	1	Y378_HALNI
23	105	3.1	775	1	VP4_ROT1
24	105	3.1	775	1	VP4_ROT1
25	105	3.1	1329	1	VP40_HUMAN
26	104.5	3.1	670	1	VG50_HSV1
27	104.5	3.1	1481	1	APV_THERT
28	104.5	3.1	2255	1	RRLV_SVS
29	104.5	3.1	2255	1	RRLV_SVS
30	103.5	3.1	1052	1	MGPC_MYCG8
31	103.5	3.1	1223	1	K125_CABEL
32	103	3.1	589	1	CPA2_PSESM
33	103	3.1	1122	1	ADP1_MYCGA

34	102.5	3.0	776	1	VP4_ROT1	P12474 bovine roca
35	102.5	3.0	2628	1	HGA_FORTI	Q51845 porphyromon
36	102	3.0	666	1	MPR2_ENTR	P39046 enterococcu
37	102	3.0	2029	1	LAR_DROME	P16621 drosophila
38	101.5	3.0	2920	1	CLR2_MOUSE	Q28130 mus muscu
39	101	3.0	1216	1	NKX1_BOVIN	P27283 simlans vir
40	101	3.0	2514	1	POLN_GINDO	P78413 homo sapien
41	100.5	3.0	519	1	TRX4_HUMAN	P40657 drosophila
42	100.5	3.0	784	1	SK15_DROME	O60231 homo sapien
43	100.5	3.0	1041	1	DD16_HUMAN	P35825 bacillus at
44	100.5	3.0	1228	1	SLAP_BACST	Q06561 caenorhabdi
45	100.5	3.0	3375	1	UN52_CABEL	

ALIGNMENTS

RESULT 1
ID VP40_HCMVA STANDARD: PRT; 708 AA.
AC P16753; Q69030;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Capsid protein P40 (Contains: Assemblin (Protease) (EC 3.4.21.97);
DE Capsid assembly protein).
GN UL80 OR APNG.
OS Human cytomegalovirus (strain AD169).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Betaherpesvirinae; Cytomegalovirus.
OX NCBI_TaxId=10360;
RN
RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF PROTEASE.
RX MEDLINE=90269039; PubMed=2161319;
RA Chee M.S., Bankier A.T., Beck S., Bohni R., Brown C.M., Cerny R.,
RA Horenell T., Hutchinson C.A., Illi, Konarsides T., Martignetti J.A.,
RA Preddie E., Satchwell S.C., Tomlinson P., Weston K.M., Barrell B.G.,
RT "Analysis of the protein-coding content of the sequence of human
RT cytomegalovirus strain AD169.";
RT Curr. Top. Microbiol. Immunol. 154:125-169(1990).
RL
RN
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF PROTEASE.
RX MEDLINE=96399136; PubMed=8805706;
RA Tong L., Qian C., Massariol M.-J., Bonneau P.R., Cordingley M.G.,
RA Lagace L.;
RT "A new serine-protease fold revealed by the crystal structure of
RT human cytomegalovirus protease.";
RL Nature 383:272-275(1996).
RN
RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF PROTEASE.
RX MEDLINE=96399136; PubMed=8805707;
RA Qiu X., Culp J.S., Dilella A.G., Hellmig B., Hoog S.S., Janson C.A.,
RA Smith W.W., Abdel-Meguid S.A.;
RT "Unique fold and active site in cytomegalovirus protease.";
RL Nature 383:275-279(1996).
RN
RP X-RAY CRYSTALLOGRAPHY (2.27 ANGSTROMS) OF PROTEASE.
RX MEDLINE=96399137; PubMed=8805708;
RA Shieh H.-S., Kurumbail R.G., Stevens A.M., Stegeman R.A.,
RA Stuman B.J., Pak J.Y., Wittwer A.J., Palmer M.O., Wiegand R.C.,
RA Howlerda B.C., Scallings W.C.;
RT "Three-dimensional structure of human cytomegalovirus protease.";
RL Nature 383:279-282(1996).
RN
RP CORE INVOLVED IN PROCESSING AND PACKAGING OF PROGENY DNA.
RX ASSEMBLIN IS A PROTEASE WHICH CAN PROTEOLYTICALLY CLEAVE ITSELF
AND THE CAPSID ASSEMBLY PROTEIN AT THE C-TERMINUS.
CC -1- CATALYTIC ACTIVITY: Cleaves -Ala-|-Ser- and -Ala-|-Ala- bonds in
CC the scaffold protein.
CC -1- PTM: CAPSID ASSEMBLY PROTEIN IS PHOSPHORYLATED.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S21.
CC -1- SIMILARITY: TO OTHER HERPESVIRUSES CAPSID PROTEIN VP40.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (<http://www.ebi.ac.uk/announcement>).
CC or send an email to license@ebi.ac.uk.

DR	EMBL; X17403; CAA53535.1; -; INIT.
DR	EMBL; X17403; CAA53535.1; ALT_INIT.
DR	PIR; S09843; Q08EB8.
DR	PDB; 1WPO; 15-OCT-97.
DR	PDB; 1LAY; 26-SEP-97.
DR	PDB; 1CMV; 04-SEP-97.
DR	PDB; 1ID4; 06-JUN-01.
DR	PDB; 1IEC; 06-JUN-01.
DR	PDB; 1IED; 06-JUN-01.
DR	PDB; 1IEF; 06-JUN-01.
DR	PDB; 1IEG; 06-JUN-01.
DR	PDB; 1JOG; 12-SEP-01.
DR	PDB; 1JUT; 12-SEP-01.
DR	PDB; 1NJT; 11-FEB-03.
DR	PDB; 1NUT; 11-FEB-03.
DR	PDB; 1NKK; 11-FEB-03.
DR	PDB; 1NKM; 11-FEB-03.
DR	MEROPS; S21.002; -.
DR	InterPro; IPR001847; Assemblin.
DR	Pfam; PF00716; Peptidase S21; 1.
DR	PRINTS; PR00236; HSYCAPSIDP40.
KM	Coat protein; Hydrolyase; Serine protease; Phosphorylation; 3D-structure.
KW	3D-structure.
FT	CHAIN 1 256 ASSEMBLIN (PROTEASE).
FT	CHAIN 257 643 CAPSID ASSEMBLY PROTEIN (BY SIMILARITY).
FT	PROPEP 644 708 C-TERMINAL PEPTIDE (BY SIMILARITY).
FT	CHAIN 1 708 GENE UL80 PROTEIN.
FT	CHAIN 336 708 GENE UL80.5 PROTEIN.
FT	CHAIN 393 708 GENE UL80.4 PROTEIN.
FT	CHAIN 478 708 GENE UL80.3 PROTEIN.
FT	SITE 256 257 CLEAVAGE (BY THE PROTEASE).
FT	SITE 643 644 CLEAVAGE (BY THE PROTEASE).
FT	ACT SITE 63 63 CHARGE RELAY SYSTEM.
FT	ACT SITE 132 132 CHARGE RELAY SYSTEM.
FT	ACT SITE 157 157 CHARGE RELAY SYSTEM.
FT	STRAND 14 23 CHARGE RELAY SYSTEM.
FT	TURN 27 28
FT	TURN 30 31
FT	STRAND 34 34
FT	HELIX 36 44
FT	STRAND 58 61
FT	TURN 62 63
FT	STRAND 64 78
FT	TURN 79 80
FT	STRAND 81 88
FT	HELIX 91 101
FT	TURN 102 103
FT	HELIX 105 109
FT	TURN 113 114
FT	HELIX 119 127
FT	STRAND 130 133
FT	STRAND 158 161
FT	TURN 167 168
FT	STRAND 170 170
FT	STRAND 172 174
FT	HELIX 177 181
FT	TURN 182 183
FT	TURN 185 186
FT	HELIX 189 197
FT	TURN 198 199
FT	STRAND 214 214
FT	HELIX 218 229
FT	TURN 230 230
FT	TURN 232 233
FT	HELIX 234 245

FT	TURN	246	246
HEIIX	247	249	
SEQUENCE	708 AA, 73851 MW, 32A993D6586824C9 CRC64;		
Query Match	3.9%; Score 130; DB 1, Length 708;		
Best Local Similarity	21.9%; Pred. No. 0.17;		
Matches	89; Conservative 42; Mismatches 151; Indels 124; Gaps 15;		
QY	269 ALPTSLPQYDVSEAVL-HTLTFRPSSAALAFVWAGLPQGTADAGTPAMEQASSG	327	
DB	319 AESPSPSPSPAPAAASHPLSAAPATA-----PPGATVAGASPA---VSS--	364	
QY	328 YLTNRHNGTTPPAGSVSYLPPEGFALERTYPNDGWTDPASAGDTTFRQVAVDEVYTN	387	
DB	365 -LAAPHGQVYLPKQAFPSLL-----	383	
QY	388 NPAGGAGAPFTVVPSPSNAYTWTFRNTLLETPSGRRLPLMP--PADFGQTVANNPK	445	
DB	384 ----GAASAVPWPYPAVA-----APPSASPPALPLPSIPASVYGAIVGYDQ	427	
QY	446 IBOGLKELT-----GCYLHSGKKNPVFQULTPASFGAVSENNPQYERTDLPD---	495	
DB	428 LAAPHADYVDPHYFGHGRYBPAPSLHPSVYPPPPS-----PAYRRRDSPEGMD	479	
QY	496 -----YGIKRSFQDN-----STVAHFRSLSHCSLYTKTYQMEG	533	
DB	480 EPPSGWERYPQGHGQSQKHRRHSGSGHNKRKETAASSSSSDBDLSFPGEAHGRAR	539	
QY	534 ---VTNNTPPGQPAHAGLLKNBERILLDLADLTRLTGYYPATQNPAAVASAFANMLSS	590	
DB	540 KRLKSHVNSDQSGSGHHG--SNQOQOQRYBELDAHEL--KRLDLPARQS--STLSLA	592	
QY	591 VLKSEATSIKISVGTAVGAAGSLAKLPGLMSVPGKTAARYA	636	
DB	593 ALPPASASSPTTTTCTPTGELTSGGGETPTALLSGAKYAEKRA	638	
RESULT 2			
MUSB_HUMAN	STANDARD; PRT; 5703 AA.		
ID_MUSB_HUMAN	Q9HCB4; 000447; 000573; 014985; 015494; 095291; 095451; Q14881;		
AC	Q99552; Q9UBZ8;		
DC	28-FEB-2003 (Rel. 41, Created)		
DT	28-FEB-2003 (Rel. 41, Last sequence update)		
DT	28-FEB-2003 (Rel. 41, Last annotation update)		
DE	Mucin 5B precursor (Mucin 5 subtyp B, tracheobronchial) (High		
DE	molecular weight salivary mucin MG1) (Sublingual gland mucin).		
GN	MUC5B OR MUC5.		
OS	Human esophagus (Human).		
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;		
OC	Mammalia; Euteria; Primates; Carnivora; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RN	SEQUENCE OF 1-1594 FROM N.A.		
RA	Chen Y., Di Y.P., Wu R.;		
RT	"Molecular cloning of the amino-terminal and 5'-flanking region of the		
RT	human MUC5B mucin gene.";		
RL	Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RN	SEQUENCE OF 1-1325 FROM N.A.		
RP	MEDLIN=99009274; PubMed=9790959;		
RA	Offner G.D., Nunes D.P., Keates A.C., Afchal N.H., Troxler R.F.;		
RT	"The amino-terminal sequence of MUC5B contains conserved		
RT	multifunctional D domains: implications for tissue-specific mucin		
RT	functions.";		
RL	Biochem. Biophys. Res. Commun. 251:350-355(1998).		
RN	[3]		
RP	SEQUENCE OF 40-1324 FROM N.A.		
RX	MEDLIN=99023932; PubMed=9804771;		
RA	Deesey J.-L., Buisine M.P., Porchet N., Aubert J.-P., Laine A.;		
RT	"Genomic organization of the human mucin gene MUC5B: cDNA and genomic		
RT	sequences upstream of the large central exon.";		
RL	J. Biol. Chem. 273:30157-30164(1998).		

[4]
RN SEQUENCE OF 1326--4895 FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=971616151; PubMed=9013550;
RA Desseyn J.-L., Guyonnet-Duperrat V., Porchet N., Aubert J.-P.,
RT Laine A.;
RN "Human mucin gene MUC5B, the 10.7 kb large central exon encodes
RT various alternate subdomains resulting in a super-repeat. Structural
RL evidence for a 1p15.5 gene family.";
RN J. Biol. Chem. 272:3168-3178(1997).
[5]
RN SEQUENCE OF 4057-4480 FROM N.A.
RP TISSUE=Salivary gland;
RC MEDLINE=97292540; PubMed=9147051;
RA Nielsen P.A., Bennett E.P., Wandall H.H., Therkildsen M.H.,
RT Hannibal J., Clausen H.;
RN "Identification of a major human high molecular weight salivary mucin
RT (MG1) as tracheobronchial mucin MUC5B.";
RL Glycobiology 7:413-419(1997).
[6]
RN SEQUENCE OF 4721-5703 FROM N.A.
RP TISSUE=Gall bladder;
RC MEDLINE=97293229; PubMed=9164870;
RA Keates A.C., Nunes D.P., Afedhal H.N., Troxler R.F., Offner G.D., C-
RT terminal sequence and genomic organization of MUC5B.";
RL Biochem. J. 324:295-303(1997).
[7]
RN SEQUENCE OF 4809-5687 FROM N.A.
RP TISSUE=Sublingual gland;
RC MEDLINE=96125355; PubMed=8554565;
RA Troxler R.F., Offner G.D., Zhang F., Iontcheva I., Oppenheim F.G.;
RT "Molecular cloning of a novel high molecular weight mucin (MG1)
RT from human sublingual gland.";
RN Biochem. Biophys. Res. Commun. 217:1112-1119(1995).
[8]
RN SEQUENCE OF 4859-5703 FROM N.A.
RP TISSUE=Placenta;
RC MEDLINE=97347489; PubMed=9201995;
RA Desseyn J.-L., Aubert J.-P., Porchet N., Laine A.;
RT "Genomic organization of the 3 region of the human MUC5B mucin.";
RN J. Biol. Chem. 272:16873-16883(1997).
[9]
RN FOLDCTION: Salivary mucin that is thought to contribute to the
RT lubricating and viscoelastic properties of whole saliva.
RL
[10]
RN SUBCELLULAR LOCATION: Secreted.
RT
[11]
RN TISSUE SPECIFICITY: Expressed mainly in bronchus glands and also
RT in submaxillary glands, endocervix, gall bladder, and pancreas.
RL
[12]
RN PTM: Highly glycosylated.
RT
[13]
RN SIMILARITY: Contains 1 TIL (Trypsin inhibitory-like) domain.
RT
[14]
RN SIMILARITY: Contains 3 VWFC domains.
RL
[15]
RN SIMILARITY: Contains 4 VWFC domains.
RT
[16]
RN SIMILARITY: Contains 1 C-terminal cysteine knot-like (CTCK) domain.
RT
[17]
RN This SWISS-PROT entry is copyright. It is produced through a collaboration
RT between the Swiss Institute of Bioinformatics and the EMBL outstation -
RL the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce>
CC or send an email to license@isb-sib.ch).
[18]
RN EMBL; AF107890; AAC33673.1; -
DR EMBL; AF086604; AAC67545.1; -
DR EMBL; AJ004863; CA06167.1; -
DR EMBL; Z72496; CA96577.1; -
DR EMBL; X74955; CA52810.1; -
DR EMBL; U63836; AAB61398.1; -
DR EMBL; U78554; AAC51344.1; -
DR EMBL; U78552; AAC51344.1; JOINED.
DR EMBL; U78553; AAC51344.1; JOINED.
DR EMBL; U78551; AAC51343.1; -
DR EMBL; U95031; AAB65151.1; -
DR EMBL; Y09788; CA470926.1; -

[illegible]

FT CONFLICT 34 34 G -> E (IN REF. 2).
 FT CONFLICT 95 100 FPGLCN -> LPELCK (IN REF. 2).
 FT CONFLICT 104 104 S -> C (IN REF. 2).
 FT CONFLICT 142 142 R -> K (IN REF. 1).
 FT CONFLICT 225 225 R -> S (IN REF. 2).
 FT CONFLICT 330 331 PL -> T (IN REF. 2).
 FT CONFLICT 337 337 E -> N (IN REF. 2).
 FT CONFLICT 336 336 E -> K (IN REF. 2).
 FT CONFLICT 362 362 G -> R (IN REF. 2).
 FT CONFLICT 369 369 MISSING (IN REF. 2 AND 3).
 FT CONFLICT 374 374 D -> N (IN REF. 2).
 FT CONFLICT 393 393 RT -> TR (IN REF. 2).
 FT CONFLICT 469 469 RK -> GR (IN REF. 2).
 FT CONFLICT 512 512 L -> P (IN REF. 2).
 FT CONFLICT 585 585 GAA -> AH (IN REF. 3).
 FT CONFLICT 601 601 A -> S (IN REF. 3).
 FT CONFLICT 628 628 DP -> RS (IN REF. 2).
 FT CONFLICT 633 633 F -> L (IN REF. 2).
 FT CONFLICT 676 676 A -> P (IN REF. 3).
 FT CONFLICT 701 701 R -> P (IN REF. 3).

Query Match 3.7%; Score 126; DB 1; Length 5703;

Best Local Similarity 22.3%; Pred. No. 6.5; Mismatches 178; Indels 118; Gaps 14;

Matches 94; Conservative 32; Mismatches 178; Indels 118; Gaps 14;

QY 221 LRPTDVPDPTEGLVTVSDYRLTY---KATCEANMPLVDGFWIGQVALTFTSLP 276
 DB 1968 LRSTATTPATSTFPAIPSSSLGTTWRLSQTTTPMAWTST-----ATPSSTP 2014
 QY 277 QYDVSAVALHLLT-----FAPSSAALA.FVMAGLPQGGAPAGTPAWMQ 322
 DB 2015 E-----TVHSTVLTTATTTGATGVAATPSTGTAHTTKVLTGTTGTAATSSSP 2067
 QY 323 ASSGGYLRHNGHTPPAGSVSYLPEGFALERYDPNDGWTDPASAGDTVTFROVAVD 382
 DB 2068 GARATLPWV-ISTTTTPTRGSGTVTSSI-----PGTHPTVLTTT 2109
 QY 383 VVVTNNPAGGSAFTTVTRVPSNATNTVFRNTLLETSPSRRLLEPMRPADPGQTVAN 442
 DB 2110 TVVATGSMATPSSSTGTPGPTLTTATTTTATGSGTTPSSSPGTPPIPVV--LTTTAT 2167
 QY 443 NPKIBOSLKEITGCVLVHSMKNPFOLETPASSFGAVSNNGYERTDLPYTGIRDS 502
 DB 2168 TPATSS-----TVPSALGTT-----HTPEVNV----- 2192
 QY 503 FDOMSTAVAHFRSLSHS-----CSIYTKYQCEGVTVNTPGQFAH-----AGILKN 552
 DB 2193 -----TTATTHGSLSPSSPHVTCTAMTSATSGILGTHITESTGSHTPAATGTGTH 2247
 QY 553 EEILCLADLALTLYGYPA-----TDNFAAASAPAA-----NMLSSVLKSEATSSIIK 602
 DB 2248 STPALSPHSSRTTSPSPGTTTPOHTATSTRTATATPSPKTRSTLPLSQPTSAFPT 2307
 QY 603 SV 604
 DB 2308 TV 2309

RESULT 3

VG93_EBV48 STANDARD; PRT; 886 AA.

AC 007284;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DE Envelope glycoprotein GP340 (Membrane antigen) (MA).
 GN BLUF1.
 OS Epstein-Barr virus (strain AG876) (Human herpesvirus 4), and
 OS Epstein-Barr virus (strain P3HR-1) (Human herpesvirus 4).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 NC NCBI_TaxID=82830, 82829;
 RN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN-AG876, and P3HR-1;
 RA MEDLINE=93331716; PubMed=8393237;
 RA Lees J.F., Arrand J.B., Pepper S.V., Stewart J.P., Mackett M.,
 RA Arrand J.R.,
 RT "The Epstein-Barr virus candidate vaccine antigen gp340/220 is highly
 RT conserved between virus types A and B."
 RL Virology 195:578-586(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-P3HR-1;
 RA Klein R., Mueller-Lantzech N.,
 RA Submitted (Oct-1992) to the EMBL/Genbank/DBJ databases.
 RL -1- FUNCTION: RESPONSIBLE FOR BBV BINDING TO THE CR2 RECEPTOR ON HUMAN
 CC B-CELLS.
 CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
 CC ENVELOPE.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).

CC EMBL, L07922; AAA02783.1; -
 DR EMBL, L07923; AAA02787.1; -
 DR EMBL, X67776; CAA47986.1; -
 DR PIR, S29605; S29605.
 DR Pfam, PF05109; Herpes BLUF1; 1.
 KM Membrane; Glycoprotein; Antigen; Late protein.
 FT CARBOHYD 47 47 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 87 87 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 195 195 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 229 229 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 328 328 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 356 356 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 378 378 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 386 386 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 435 435 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 443 443 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 457 457 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 497 497 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 519 519 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 533 533 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 568 568 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 589 589 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 624 624 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 680 680 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 886 AA; 92388 MW; 4394F6130DECCA8A CRC64;

Query Match 3.6%; Score 123; DB 1; Length 886;

Best Local Similarity 20.3%; Pred. No. 0.73; Mismatches 269; Indels 206; Gaps 29;

Matches 138; Conservative 66; Mismatches 269; Indels 206; Gaps 29;

QY 48 NPTAAADLAQSDANTVTPPANISMPFERNAKRIIDSDSIGYFKYLDPAQATES 107
 DB 169 NITAVVR--AQGD--VTLPLSLPTSAQDSNPS-----VKTELG--NEID-----IEC 211
 QY 108 ARAVGSXKIPDGLVPSVDATIRE-----IYNBSCPVTVDVSLDGRKMSLSISF 160
 DB 212 IMEDGISOVLPEDNKFNITCSGYSHVPSGGILTSTSPAT--PIGTYAASLKRTPR 269
 QY 161 PMFR-----TAYVAANVENKEMSLD-----VYNDLI-----EWLNINLADWRYYVDS 203

```

Db 270 PVSRFLGNSILYFYSGNGPKASGDYCIOSNIWFSDEIPASQDMPNTTIDITYVGN- 328
Qy 204 QMINFTDITYYRIRVLRPTVDPPBGLVRTVSDYRLTYKAITCEANMPTLVDOGF- 263
Db 329 -----ATYSVPMVTS-----DANSPNVTVAFW 352
Qy 264 -----IGQYALPT----- 273
Db 353 AMPNNTETDFCKMTLISGTSGCENISGAFASNRFPDITVSGLGTAETKILITRTATNA 412
Qy 274 -----SLPQYDVSEAYALHTLTPAPSSAALAFVWAGLPGGTAPAGTPAMEQAS 324
Db 413 TTTTHXVIFSKAPSESTTSPILNTTGFAPNT-----TGLBSSTHVPNTLVA--PAS 463
Qy 325 SGGLYTWKHNHTTPPAGSVSYLPEGFALERYDPNDG-----SWTDFASAGDTVTRQ 377
Db 464 TGPVSTADVTSPAGTTSASP---VTPSPSPDNQTESKAPDMTSPSAVTPTPNA 520
Qy 378 VAVDEVVVT-----NPNAGGSAPTFTYRVPPSNAYMT-----VFNTLLETRPS 423
Db 521 TSPPTAVTTPPNATSPTLGKTSPTSAVTPTPNATSPTPAVTTPPNATPTLGKTSPT 580
Qy 424 SRLLELPMPPADFGQTVANMPKIEOSLKEITLGCYLVSQKRNIPVQLTTPASSFGAVSFN 483
Db 581 S-ATTTTPPNATSPVGETSP--QANTTNHTLG-----GTSSTPVVTSPPKATSAVTTG 632
Qy 484 NPGYERTRDLPDYTGIR-DSFPDQNSTAVAHFRSLSHCSIVTKTYQMEGVTVNTPFG 542
Db 633 Q--HNITSSSTSSMSLRPSISSETLSPSTD--NSTSHMPLLTLSAHPGCEMITQV--TPAS 688
Qy 543 QFALAGLKLKEIILCLADLTLRLTGVPAT-----DNRAAVASAAAMLSVLKSE 595
Db 689 TSTH-----HVSSTSPAPRGTTSGAGSGNSSTKPGEBVNTVTKGTPKN 734
Qy 596 ATSSIIKSGEATVCAQAS 614
Db 735 ATSPQABSGOKTAVPTVTS 753

RESULT 4
VGP3_EBV STANDARD; PRT; 907 AA.
ID_VGP3_EBV
AC P03200; P03201;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DE 28-FEB-2003 (Rel. 41, Last annotation update)
DR Envelope glycoprotein GP340/GP220 (Membrane antigen) (MA).
GN BLUF1.
OS Epstein-Barr virus (strain B95-8) (Human herpesvirus 4).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirinae.
OX NCBI_TaxID=10377;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS GP220 AND GP340).
RX MEDLINE=84270667; PubMed=6087149; Deisinger P.L., Farrell P.J.,
RA Baer R., Bankier A.T., Biggin M.D., Deininger P.L., Farrell P.J.,
RA Gibson T.J., Hatfull G., Hudson G.S., Satchwell S.C., Seguin C.,
RA Tuffnell P.S., Barrett B.G.;
RT "DNA sequence and expression of the B95-8 Epstein-Barr virus genome.";
RL Nature 310:207-211(1984)
CC -1- FUNCTION: RESPONSIBLE FOR EBV BINDING TO THE CR2 RECEPTOR ON HUMAN
CC -1- B-CELLS.
CC -1- SUBCELLULAR LOCATION: MOST ABUNDANT COMPONENT OF THE VIRAL
CC ENVELOPE.
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Name=GP340;
CC IsoId=P03200-1; Sequence=Displayed;
CC Name=GP220;
CC IsoId=P03200-2; Sequence=VSP 002070;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; V01555; CAA24854.1; -.
DR PIR; A43042; Q08B21.
DR PIR; B43042; Q08B22.
DR Pfam; PF05109; Herpes_BLP1; 1.
KW Membrane; Glycoprotein; Antigen; Late protein; Alternative splicing.
FT CARBOHYD 47 47
FT CARBOHYD 87 87
FT CARBOHYD 114 114
FT CARBOHYD 166 166
FT CARBOHYD 169 169
FT CARBOHYD 195 195
FT CARBOHYD 229 229
FT CARBOHYD 277 277
FT CARBOHYD 318 318
FT CARBOHYD 328 328
FT CARBOHYD 345 345
FT CARBOHYD 356 356
FT CARBOHYD 378 378
FT CARBOHYD 386 386
FT CARBOHYD 411 411
FT CARBOHYD 435 435
FT CARBOHYD 443 443
FT CARBOHYD 457 457
FT CARBOHYD 497 497
FT CARBOHYD 519 519
FT CARBOHYD 533 533
FT CARBOHYD 547 547
FT CARBOHYD 568 568
FT CARBOHYD 589 589
FT CARBOHYD 610 610
FT CARBOHYD 624 624
FT CARBOHYD 627 627
FT CARBOHYD 645 645
FT CARBOHYD 656 656
FT CARBOHYD 683 683
FT CARBOHYD 701 701
FT CARBOHYD 735 735
FT CARBOHYD 746 746
FT CARBOHYD 755 755
FT CARBOHYD 780 780
FT CARBOHYD 815 815
FT CARBOHYD 858 858
FT CARBOHYD 888 888
FT VARSPLIC 502 698
SQ SEQUENCE 907 AA; 94431 MW; 0750141CBAC5269 CRC64;
Query Match 3.6%; Score 121.5; DB 1; Length 907;
Best local similarity 19.6%; Pred. No. 0.96; Index 233; Gaps 28;
Matches 138; Conservative 66; Mismatches 266;
48 NPTAADLQSLDANTVTPPNATISMPDEFNMAKGLIDSDSIGWTFKIDPAGATES 107
Db 169 NITAVVR--AQGLD---VTLPISLPTGAQDSNPS-----VTNMLG--NRID----IBC 211
Qy 108 ARAVGEYSKIPDGLVQSVDAIRE-----IYNECPVTVDSVPLDGRQWSLSPF 160
Db 212 IMBGRISQVLPQGNKRNITCGYSHVPSGGILITSPVAT--PIPTGYAVSLRLTPR 269
Qy 161 PMPR-----TAYAVAVNENKMSLD-----VNDLI-----EMLNLDLMVYVYVSE 203
Db 270 PVSRFLGNSILYFYSGNGPKASGDYCIOSNIWFSDEIPASQDMPNTTIDITYVGN- 328
Qy 204 QMINFTDITYYRIRVLRPTVDPPBGLVRTVSDYRLTYKAITCEANMPTLVDOGF- 263
Db 329 -----ATYSVPMVTS-----DANSPNVTVAFW 352

```

```

QY 264 -----IGQVATLP----- 273
DB 353 AMPNNTETDPKCKMTLTSGTSPSCENISGAFASNFRTDITVSGTAGTAPKTLITRTATNA 412
QY 274 -----SLPOYDSEAVYALHTLTFARSSAALAFWAGLPGCGTAPAGPAMEQAS 324
DB 413 TTTTHKVIYSKAPESSTISPTLTGFPADPNTT-----TGLPSTHVPMLTA--PAS 463
QY 325 SCGYLWRHNGTTPFAGSVSYVLPEGFALERYPDNDGWS-----TDFASAGDTV 373
DB 464 TGPVTSTADVTSPFAGTTSASGASP-----VTPSSPMDNGTESKAPMTSSTSPVTTP 516
QY 374 TTRQAVAVDEVVT-----NNPAGGSAFTFTRVPSNAYNTVTR 414
DB 517 TPNATSPPTAVTTPPTNATSPPTAVTTPPTNATSPPTGTSPTSAVTTPTPTNATSPPTLCK 576
QY 415 N-----TLETPSSRLLELMPRPDQGTAYANNKIGOSLKEITGCV 458
DB 577 TSPSTAVTTPPTNATSPPTGTSPTSAVTTPTNAT--GPTVGETSP--QANNATNHTLG-- 631
QY 459 LVHAKRNPFVQLTTPASSFGAVSFNNPGYERTDLDPYTGIRDSFDQNNSTAVAFRSLS 518
DB 632 ---GTSPTFVTSQPKNATSAVTTGQ--HNITSSSTSSMSLRSSNPETLSPTSDNSTS 686
QY 519 HSCSIYTKTYQMGEGVTNNTPPGQFAHAGLKNBEILCLADDLATRLTGVPAT----- 573
DB 687 HMPLLTSAHPTGENTITOV--TPASISTH-----HVSSTSPAPRPCTTSQAS 731
QY 574 --DNFAAASAFANMLLSVLKSEATSIKISGETAVGAQS 614
DB 732 GGNSTSTKPGENVNTKGTTPONATSPQAPSGQKTAVPTVTS 774

RESULT 5
ILPR_BRAVA
ID ILPR_BRAVA STANDARD; PRT; 1363 AA.
AC 002466;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Inulin-like peptide receptor precursor (BC 2.7.1.112) (ILP receptor).
OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_Taxid=7740;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96408719; PubMed=8813726;
RA Pashmforoush M., Chan S.-J., Steiner D.F.;
RT "Structure and expression of the inulin-like peptide receptor from
RT amphioxus."
RL Mol. Endocrinol. 10:857-866(1996).
CC - FUNCTION: THIS RECEPTOR BINDS TO THE INSULIN RELATED PEPTIDE AND
CC - HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
CC - CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
CC tyrosine phosphate.
CC - SUBUNIT: PROBABLE Tetramer of 2 alpha and 2 beta chains linked by
CC disulfide bonds. The alpha chains contribute to the formation of
CC the ligand-binding domain, while the beta chain carry the kinase
CC domain (by similarity).
CC - SUBCELLULAR LOCATION: Type I membrane protein (by similarity).
CC - SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
CC RECEPTOR SUBFAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; S83394; AAB50848.1; -

```

```

DR FIR; T43220, T43220.
DR HSP, P06213, IIRK.
DR InterPro; IPR000494; EGFR_L_domain.
DR InterPro; IPR003961; FN III.
DR InterPro; IPR006211; Furin-like.
DR InterPro; IPR006212; Furin repeat.
DR InterPro; IPR000719; Prot_Kinase.
DR InterPro; IPR002011; RTKinaseII.
DR InterPro; IPR001245; Tyr_kinase.
DR Pfam; PF00041; fn3; 3.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; pkinase; 1.
DR Pfam; PF01030; Recept_L_domain; 2.
DR PRINTS; PR00109; TYRKINSE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00261; FU; 3.
DR SMART; SM00219; TYRK; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transferase; Tyrosine-protein kinase; Receptor; Transmembrane;
KW Glycoprotein; ATP-binding; Phosphorylation; Signal.
FT SIGNAL 1 29
FT CHAIN 30 716
FT FT
FT PROPEP 717 720
FT CHAIN 721 1363
FT FT
FT DOMAIN 721 928
FT TRANSMEM 929 949
FT FT
FT DOMAIN 950 1363
FT NP_BIND 994 1283
FT BINDING 1000 1008
FT ACT_SITE 1028 1028
FT MOD_RES 1148 1148
FT MOD_RES 1174 1174
FT CARBOHYD 51 51
FT CARBOHYD 97 97
FT CARBOHYD 137 137
FT CARBOHYD 278 278
FT CARBOHYD 463 483
FT CARBOHYD 599 599
FT CARBOHYD 617 617
FT CARBOHYD 665 665
FT CARBOHYD 666 666
FT CARBOHYD 711 711
FT CARBOHYD 712 732
FT CARBOHYD 736 736
FT CARBOHYD 743 743
FT CARBOHYD 816 816
FT CARBOHYD 885 885
FT CARBOHYD 898 898
SQ SEQUENCE 1363 AA; 154104 MW; 23812084EAB1D065 CRC64;

Query Match 3.5%; Score 119; DB 1; Length 1363;
Best Local Similarity 19.8%; Pred. No. 2.6;
Matches 99; Conservativity 60; Mismatches 164; Indels 180; Gaps 23;

QY 79 NMAKIKIDSDSISIMYRK-----YLDPAQATSAARAVGYSIKIPGLVYKPSVDA 128
DB 416 DMDRTDITIDBGKLFHPKPCRHVILTMVDKGLPERA-----ITD----- 458
QY 129 EIREIYNBCEPVTVDSVPLDGRQKSLIFSPMRTAYVAVYENKEMSLDVVNDLIE 188
DB 459 -----TIDSTLWNGDQ-----AQCSRSRIE-----IE 480
QY 189 WLNALADRYVDSHOWINFTNDTYVYRIRVLRPTVDVDPPTBGLVATVSDYRLTYKAI 248
DB 481 EINTSKDMITL-----RWSEF-----PDRDLSTVVS--YRET----- 514
QY 249 TCEANPMLTVDOGF-WIGQVATLPVTS-LPOYDSEAVYALHTLTFARSSAAL-----ATFW 304

```

```

Db 515 -----EDQIDEVGDQACGNTWKERDVSPYQAHIIITGLKPTQYALLKVTYK 565
Qy 305 AGLPQGG-----TRPACTPAMEQASGGLYLTWRNGTTTPAGSYVLP 348
Db 566 AGAREGSGAKSDIYVARTADKPHIPQDVVYNSNLTIIITWKPNN--PRGVNTYIV 623
Qy 349 EGPALEPYDPNDGWSWTFASAG-----DTVTFROVADEVVYVNNPAGGSAFTTVVP 403
Db 624 K-YRQQEDVAMEQREYCKGGLKPHRPTQGLIEDIVNNEEPNNSTIDGTC-----CECP 678
Qy 404 PS-----NAYTTFVRNTLLETPSSRLLELPPPADP--GGTV--- 440
Db 679 KSEDEIRIEBEAARQGEFENFLHNHYHKKENETRARRRRELVTARPFYSNQTVAVT 738
Qy 441 -----ANNKIEOSLKETLGGCYLVHSKRN--PVPQULTPSSFGAVSFN 483
Db 739 LPSTNRTPPTPTPNPNQLETTVNNHNM--VVLTGLRHFSYIIEVIACNADAAVGS 795
Qy 484 NPGYERTDLPDYTGIRDSFD 504
Db 796 GSAVELARTQAD-----DSAD 811

RESULT 6
CH11_BACCI
ID CH11_BACCI STANDARD; PRT: 699 AA.
AC P20533;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Chitinase A1 precursor (EC 3.2.1.14).
GN CH11.
OS Bacillus circulans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1397;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=WL-12;
RX MEDLINE=90368776; PubMed=2203782;
RA Watanabe T., Suzuki K., Oyanagi W., Ohnishi K., Tanaka H.;
RT "Gene cloning of chitinase A1 from Bacillus circulans WL-12 revealed
RT its evolutionary relationship to Serratia chitinase and to the type
RT III homology units of fibronectin";
RL J. Biol. Chem. 265:15659-15665(1990).
RN [2]
RP MUTAGENESIS.
RC STRAIN=WL-12;
RX MEDLINE=93366760; PubMed=8103047;
RA Watanabe T., Kohori K., Miyashita K., Fujii T., Sakai H.;
RT "Identification of glutamic acid 204 and aspartic acid 200 in
RT chitinase A1 of Bacillus circulans WL-12 as essential residues for
RT chitinase activity";
RL J. Biol. Chem. 268:18567-18572(1993).
CC -I- CATALYTIC ACTIVITY: Hydrolysis of the 1,4-beta-linkages of N-
CC acetyl-D-glucosamine polymers of chitin.
CC -I- SIMILARITY: Contains 2 fibronectin type III domains.
CC -I- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M57601; AAA81528.1; -.
CC PIR: A38368; A38368.
CC PDB: 1BD7; 24-MAY-00.
CC PDB: 1K85; 18-DEC-02.

```

```

DR InterPro: IPR003610; CBM 5.12.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR003962; FNIII subd.
DR InterPro: IPR001223; Glyco_hydro_18.
DR InterPro: IPR001579; Glyco_hydro_18/2.
DR Pfam: PF02839; CBM_5_12; 1.
DR Pfam: PF00041; fn3; 2.
DR Pfam: PF00704; Glyco_hydro_18; 1.
DR PRINTS: PR00014; FNTYPRIII.
DR PRINTS: PD000471; Glyco_hydro_18; 2.
DR SMART: SM00495; ChEBD3; 1.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00636; Glyco_18; 1.
DR PROSITE: PS01095; CHITINASE 18; 1.
KW Hydrolyase; Glycosidase; Chitin degradation; Signal; Repeat;
KM 3D-structure.
FT SIGNAL 1 41
FT CHAIN 42 699
FT DOMAIN 42 460
FT DOMAIN 465 549
FT DOMAIN 560 644
FT ACT_SITE 204 204
FT ACT_SITE 200 200
FT MUTAGEN 200 200
FT MUTAGEN 204 204
SQ SEQUENCE 699 AA; 73677 MW; AC7C9B2282987643 CRC64;

Query Match 3.4%; Score 115.5; DB 1; Length 699;
Best Local Similarity 21.4%; Pred. No. 1.8;
Matches 114; Conservative 54; Mismatches 207; Indels 157; Gaps 25;

Qy 195 DMRVV-----DSEMINFNDTYYVRIYVAPTDVPTGIVRTVDYLYYAI 248
Db 202 DWETPVSGGLDGNKRPBKQNTYLLSKIREKDAAGAVGKYLITAGASATYAA 261
Qy 249 TCEANMPTLVDOGFMIGQVALTPTSLOYVSHAY--ALHTLTPAPSSAALAFYWA 305
Db 262 TELKIAIYD--WI-----NIMTFDNGAMKISAHNAPLINDPASA--A 304
Qy 306 GLPQGGT-----APACTPA-----WE--QASGGYLTWRNGT--T 337
Db 305 GVPDANTEFNVAAGAGHLDAGVPAKVLGVPPYGRGWDGCAQAGNGQYCTCGSSVGT 364
Qy 338 PPAGS-----VSYVLEPGALERYDPNDGSWD-----PASAGT---- 372
Db 365 WEAGSPFDYIEANYINKNGYT--RY-----WMDTAKVPYLYNASNRRFISYDAESVG 416
Qy 373 -----VTFROVADEVVYVNN-----PAGGGSAPFTTVVP--PSNAYTNTV 412
Db 417 YKTAYIKSKGLGAMFHELSGDBRKTLQNKALKALPTGGTYPVDDTTAPSVGNARSTGV 476
Qy 413 FRNTL-LETRPSSRLLELPPPADPFGQTVANNPKIEOSLKEYL-----GCVLVHSK 463
Db 477 TANSVTLAMNASTNVGV-----GYVVYGANLATSVTGTATISGLTAGSTYPTIK 530
Qy 464 MENVVPQUTPASPAGVSPNNPGYERTRDLPDYTGIDSPQGNSTAVAHRSLSHSGSI 523
Db 531 AKDAAGNISASNAVTVS-----TTAPQGGTQAPTAFTNLAFTAQTTSS 575
Qy 524 VTKTYQGWEGTANNTPFGQFAHAGLLKNEEILCLADLITRLGVVYPATONPAAVS-A 582
Db 576 ITLS--WTASTDVGCTGVYVNGTA-----LATVIGTTATISGLAADTSYT 621
Qy 583 FAANMLSSVLNKEATSSI--IKSVGET--AVGAAGSLATLPGILMSVPGK 629
Db 622 FTVAKDAAGVSAASNAVSVKTAFTNPGVSAWQVNTAVTAQGLVTVNCK 673

RESULT 7
ALAI_CANAL
ID ALAI_CANAL STANDARD; PRT: 1419 AA.
AC O13368;
DT 15-JUL-1999 (Rel. 38, Created)

```

DT 15-JUL-1999 (Rel. 38, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Agglutinin-like protein ALAI precursor (Agglutinin-like adhesin).
 GN ALAI OR ALIS.
 OS Candida albicans (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 CC Saccharomycetales; mitospotic Saccharomycetales; Candida.
 RX NCBI_Taxid=5476;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=98053977; PubMed=9339828;
 RA Gaur N.K., Klotz S.A.;
 RT "Expression, cloning, and characterization of a Candida albicans
 gene, ALAI, that confers adherence properties upon Saccharomyces
 cerevisiae for extracellular matrix proteins."
 RT Infect. Immun. 65:5289-5294(1997).
 RL -1- FUNCTION: MAY PLAY A ROLE IN ADHESION AND PATHOGENESIS.
 CC -1- PTM: N-GLYCOSYLATED AND O-GLYCOSYLATED (POTENTIAL).
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; AF025429; AAB88883.1; -
 DR Cell adhesion; Glycoprotein; Repeat; Signal.
 DR PIR; T30531; T30531.
 KM
 FT SIGNAL 1 17
 FT CHAIN 18 1419 AGGLUTININ-LIKE PROTEIN ALAI.
 FT DOMAIN 399 404 POLY-THR.
 FT DOMAIN 408 418 POLY-THR.
 FT DOMAIN 437 441 POLY-THR.
 FT DOMAIN 673 676 POLY-SER.
 FT DOMAIN 687 690 POLY-SER.
 FT DOMAIN 700 703 POLY-SER.
 FT DOMAIN 719 724 POLY-SER.
 FT DOMAIN 749 752 POLY-SER.
 FT DOMAIN 787 791 POLY-SER.
 FT DOMAIN 869 872 POLY-SER.
 FT DOMAIN 875 883 POLY-SER.
 FT DOMAIN 901 911 POLY-SER.
 FT DOMAIN 1216 1221 POLY-SER.
 FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 919 919 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1301 1301 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1326 1326 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 1419 AA; 149635 MW; 249F33F688A9D586 CRC64;
 Query Match 3.4%; Score 115; DB 1; Length 1419;
 Best Local Similarity 17.7%; Pred. No. 5.3;
 Matches 123; Conservative 90; Mismatches 249; Indels 232; Gaps 26;
 QY 39 TGRQVPPPN--FTAAADLAOSLDNATYTPPANISSMPEFRNMAKGD----- 86
 DB 139 TGSVDLEDSKCTAG-----TNYTFNDGSKSLAVAFKSTVDQSGYLTTSR 189
 QY 87 -----LDSISGMFYKLDPA----- 102
 DB 190 MSLNKIATLYVAPOCENGYSSTMGSTSYSDVALDCSNVHIGSKGVNDMHPYTS 249
 QY 103 -GATESARAVG--EYSKIPDGLVKSVDAREIYNEBCPVVTVDSVPLDGRQMSLIF 158
 DB 250 FSYTKSCSSFGISITQONVPAGRPF-IDAYIS-----PSDNNQYQLS-- 291
 QY 159 SFPMPTAYAVANVENKESLDVNDLEWLNLDMDRYVDSQGINFTDPTTYVRI 218
 DB 292 ----YKNDYTCVDY-----WQAPPTLTKWTGYKNSDAGSNGI 325
 QY 219 RVLRFYVDPDPTEGLVRTVSDVRLTYKAITCE--ANMPLTVQGFWIG----- 265

DB 326 VIVATRTVDSSTA-VTLLPFPNSVDKTKTIBLQPIPTTTITTSYGVVSTSTKAP 384
 QY 266 -GQYALPTSLPYD-----VSEAYALHTLPARSSAALAFVAGLPQGTAPAGPAW 320
 DB 385 IGRATVAVDPVPTTITTVVSEWGTITTTTRNPNPDSIDTVVQVPSNPPTTITQFW 444
 QY 321 HQASGGVLTWRHNGTTPAGSVSVLPEGPALERYDNDGS-----MTDPASAGPVT 374
 DB 445 SES-----FTSTTITNSLKGTDVIVRE-----PNNFTVTFEFSRPAITETIT 491
 QY 375 PQYAVDEVVV-----TNNPAGSGAPFTVAVPPSNAYTN 410
 DB 492 SKPGTGSVIVREPHNPVTTTTFEFSYSATETITNGPBGTDV---IVREPHNPVTT 548
 QY 411 TVPNTLLTTPPSRRLLEPMPADPQGTANPK-IEQSLKLTGLCTLVHSGRRNVF 469
 DB 549 TKFMSSEYALT-----ETINKKEGTDVIVKEPYNPVTTFEFSRSY 592
 QY 470 QLTASRFG-----AVSFNNPGYERTDL-----PDVTGIRDSPO 505
 DB 593 ATTEITNGPGTGSVIVREPHNPVTTTTFEFSYSATETITNGPBGTDVIVHDPLE 652
 QY 506 NMSTAVAFRSLSHSCSIVTQYQWBGVTNVTTPQGFANAGLAKNEILCLADLADR 565
 DB 653 SSSIT-----AISSDSMISSAQ--BSSSVBQ---SSSIVGLSSSDIPLSDMPSS 702
 QY 566 LTGYVPATDNFAAVSAPANMSSVLKSEATSS 599
 DB 703 STGL-----TSSESGTVSYSDSSSIBSSTLSS 733

RESULT 8
 COAT_PAVHB STANDARD; PRT; 781 AA.
 ID COAT_PAVHB
 AC P07299;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-APR-1988 (Rel. 07, Last sequence update)
 DT 01-APR-1990 (Rel. 14, Last annotation update)
 DE Probable coat protein VP1.
 OS Human parvovirus B19.
 OC Viruses; ssDNA viruses; Parvoviridae; Parvovirinae; Erythrovirus.
 RX NCBI_Taxid=10798;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Isolate AU.
 RX MEDLINE=86200451; PubMed=3701931;
 RA Shade R.O., Bundell M.C., Colmore S.F., Tattersall P., Astell C.R.;
 RT "Nucleotide sequence and genome organization of human parvovirus B19
 isolated from the serum of a child during aplastic crisis."
 RL J. Virol. 58:921-936(1986).
 CC -1- SIMILARITY: BELONGS TO THE PARVOVIRUSES COAT PROTEIN FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; M13178; AAA66867.1; -
 DR PIR; A24299; VCPV19.
 DR InterPro; IPR001403; Parvo_coat.
 DR Pfam; PF00740; Parvo_coat; 1.
 KM Coat protein; Glycoprotein.
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 184 184 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 220 220 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 259 259 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 781 AA; 86015 MW; 8C6254DBD0576B07 CRC64;
 Query Match 3.4%; Score 113.5; DB 1; Length 781;
 Best Local Similarity 19.5%; Pred. No. 2.9;

Matches	128;	Conservative	51;	Mismatches	191;	Indels	287;	Gaps	28
OY	112	GEYSKIPRGGLYKFSYDAIR---	EYNBECVTVDSVPLDGRQMSISFSPMFRAY	167					
Db	85	GQSDHPHALSSSSSHASPRGENAVLSSSEHLHKPCQVSVQLPG-	-----TWY	130					
OY	168	VAVANVEKMSLDVVDNLLIEMLNLADRY-----	VYDSBQIMPTNDT	212					
Db	131	VGPGR-----ELQAGPQSAVDSAAIRHIDRRYSQALKGINPYTHMTVADSEELKKIKNET	186						
OY	213	TYVVAVIRLPRPYDVDPDEGLVTRTSDRLTYKAITCEANNPTLVDQFMIGQYALTP	272						
Db	187	GFGQAV-----VKQY-FLTKG---AARVAHFQ-----SL	213						
OY	273	TSLPOYDVSAYVALHTLTFARPSSAALAFAVMAGLPQGGTAPAGTPAMEQASSGGYLTWR	332						
Db	214	PEVPAYNASEKPYSMWT-----SVNSAEASTGAG--GGGSNSVKSMSB-	254						
OY	333	HNGTFFPAGSVSYVLREGFALERYP-----	358						
Db	255	--GATFASNSVCTCFSRQFLIP-YDPEHHYKVPSPASSCHNASGKEAKVCTISPIMGYS	311						
OY	359	-----NDGSWTDVFAAGDPVTFEROYAVDEVVTTNNPAGGG	393						
Db	312	TPWKYLDNALNLFFSPLEFQHLINENSS---IADALTVTISLAIVDV---TDKTTGGG	365						
OY	394	-----SAPFTVARV--PSPNAYTNTVFRNTLET	420						
Db	366	VQVTDSTTGRLCMLVDHEKYPYVLAGQGDTLAPELPIVWPYPPOYALVTVGDVNT----	421						
OY	421	RPSRRLELPMRPADPGQTVANNPKIEGSLKETLACY-LVNSKRKNRPVFLQTPASSFGA	479						
Db	422	-----QGISGDSK--KLAEBSAPVYLEHSS--FOLLGTGGTAS	456						
OY	480	VSFN-----NPGYERTDLPDYTGIRDSPQDNSTAVAHFRLSH	519						
Db	457	MSYKRPYPPEVLEEGSCQHFYMTNPLVGSRLGVPDTLG-----GPKRSLTH	505						
OY	520	-SCSIVTKTYQGMEGVTNVNTPFGQFAHAGLKNBEILCLADDLATRLGVVPA-----	572						
Db	506	EDHAIQPNFMGPVLNVSSTVEGSSNNGAKKALTGSTGNSQNTNIS-LAPGPVSCP	564						
OY	573	---TDNFAAAYSAPAAAMLSSVLKSEATSIIKSVGFATVGA-----QSLGTLAP	620						
Db	565	HHMDTKDYVTGINAI-----SHGQTTYNAEDKEXQVQGVGRFP	602						
RESULT 9									
SACB_STRSL STANDARD; PRT; 969 AA.									
AC	055242;								
DT	15-DEC-1998 (Rel. 37, Last sequence update)								
DT	15-DEC-1998 (Rel. 37, Last annotation update)								
DT	15-DEC-1998 (Rel. 37, Last annotation update)								
DE	Levanucrase precursor (EC 2.4.1.10) (Beta-D-fructofuranosyl								
DE	transferase) (Sucrose 6-fructosyl transferase).								
GN	FTF.								
OS	Streptococcus salivarius.								
OC	Bacteria; Firmicutes; Lactobacillales; Streptococcaceae;								
OC	Streptococcus.								
OX	NCBI_TaxID=11304;								
RN	SEQUENCE FROM N.A.								
RC	STRAIN=ATCC 25975;								
RX	MEDLINE=93322332; PubMed=8331080;								
RA	Ratcliam C., Giffard P.M., Jacques N.A.;								
RT	"The cell-bound fructosyltransferase of Streptococcus salivarius: the								
RT	carboxyl terminus specifies attachment in a Streptococcus gordonii								
RT	model system."								
RL	J. Bacteriol. 175:4520-4527 (1993).								
CC	- CARBOLYTIC ACTIVITY: Sucrose + ((2,6)-beta-D-fructosyl)(N) =								
CC	- glucose + ((2,6)-beta-D-fructosyl)(N+1).								
CC	- SUBCELLULAR LOCATION: SECRETED. CELL-WALL BOUND.								

CC -1- SIMILARITY: BELONGS TO FAMILY 68 OF GLYCOSYL HYDROLASES.

CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.1sb-sdb.ch/announce>

CC or send an email to license@1sb-sdb.ch).

CC -----

DR EMBL, L08445; AAA71925.1; -

DR InterPro: IPR003469; Glyco_hydro_68.

DR InterPro: IPR005877; Gpos_Ysirk_

DR Pfam; PF0435; Glyco_hydro_68; 1.

DR Pfam; PF04650; Ysirk_signal; 1.

DR TIGRPFams; TIGR01168; Ysirk_signal; 1.

KM Transferase; Glycosyltransferase; Signal; Cell wall.

FT SIGNAL

FT CHAIN

FT ? 969 LEVANSUCRASE.

FT SEQUENCE 969 AA, 103983 MW, D3895B32AC7735A CRC64;

Query Match 3.3%; Score 113; DB 1; Length 969;

Best Local Similarity 21.0%; Pred. No. 4.2;

Matches 81; Conservative 47; Mismatches 141; Indels 116; Gaps 17

QY 97 KYLDPAQTSAARAVGEYSKIPDGLVRFVSVAIREIYNBSCPVTVDVSPIDGRQWSLS 156

DB 568 KSTDAEGTVAAAREAVGD-DVVMLGFSVSDSLRGKXRLPLNGS--GVVLTASVPADMTSTYS 624

QY 157 IESPP-----MRTAV---AVANENK-----EMSLDVNVDLEMLNLADM 196

DB 625 YVAAPFVBSPTLTVSTINRGGIAGAKNSYWAAPSFLIKNNADDTTEVLEKMTNQGM 684

QY 197 RYVVDSQWIMFTNDTTYVIRVRLPYVDYDPDTBGLVRTVSDYRLTYKAITCEANMPT 256

DB 685 IMDSSBSLVHVGQNSAKL-----PNE-----DENVDYAVS----- 717

QY 257 LVDOGFWIGGYALTPSISLPQYD--VSRVYALHTLTTPARPSSAALAFWAGLPQGGT 312

DB 718 -----GYGLKPHYPTVDSGTGSEAHGVLTVA-----VQDGD 751

QY 313 APAGTPARQASSGCYLTWRN-----CTTPPA-GSVSYVLPREGALERTYDNDGSM 363

DB 752 KKADPPTPVSPTRG---NHSVDKTKPKPTGSRKADNNQPSADKEDKPTNPTNDSPAR 807

QY 364 TDFPASAGDTVFRQVAVD-EVYVNTNPAQGS-----APFTVAVPS-----NA 407

DB 808 TPFPTGYGSHSDNNSSNDHVAVPKPGTSGSVGDRFRVAQAARATVPEPKTIYATGPV 867

QY 408 YTNVTFRNTLTETRPSSRLRLPMP 432

DB 868 PTNVTKEBSVTET-----EAPKR 885

RESULT 10

GUXB_CELFI

ID GUXB_CELFI STANDARD; PRT; 1090 AA.

AC P50859;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-OCT-1996 (Rel. 34, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Exoglucanase B precursor (BC 3.2.1.91) (Exocellulobiohydrolase B)

DE (1,4-beta-cellobiohydrolase B) (CBP120).

GN CBBH OR CREB.

OS Cellulomonas fimi.

OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;

OC Micrococciaceae; Cellulomonadaceae; Cellulomonas.

OX NCBI_Taxid=1708;

RA [1]

RP SEQUENCE FROM N.A., AND SEQUENCE OF 456-461.

RC STRAIN-ATCC 484;

RC MEDLINE=96003898; PubMed=7575482;

RA Shen H., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;

RT "Cellulohydrolase B, a second exo-cellulohydrolase from the
 RT cellulytic bacterium Cellulomonas fimi.",
 RL Biochem. J. 311:67-74(1995).
 RN [2]
 RX SEQUENCE OF 54-75. Pubmed=8458833;
 RA MEDLINE=93209933; PubMed=8458833;
 RT Meinke A., Gilkes N.R., Kilburn D.G., Miller R.C. Jr., Warren R.A.J.;
 "Cellulose-binding polypeptides from Cellulomonas fimi: endoglucanase
 RT D (Cend), a family A beta-1,4-glucanase.",
 RL J. Bacteriol. 175:1910-1918(1993).
 RN [3]
 RP SEQUENCE OF 54-78. Pubmed=8147863;
 RX MEDLINE=94197708; PubMed=8147863;
 RA Shen H., Tomme P., Meinke A., Gilkes N.R., Kilburn D.G.,
 RT Warren R.A.J., Miller R.C. Jr.;
 "Stereocatalytic course of hydrolysis catalysed by Cellulomonas fimi
 RT Cend, a member of a new family of beta-1,4-glucanases.",
 RL Biochem. Biophys. Res. Commun. 199:1223-1228(1994).
 CC -1- FUNCTION: Hydrolyze cellobiose to a mixture of cellobiose,
 CC cellobiose and cellobiose, with only a trace of glucose. It
 CC hydrolyzed cellobiose to cellobiose and cellobiose, and
 CC cellobiose to cellobiose, but it did not hydrolyze cellobiose.
 CC Has also weak endoglucanase activity. Hydrolyzes glucosidic bonds
 CC with inversion of anomeric configuration.
 CC -1- CATALYTIC ACTIVITY: Hydrolysis of 1,4-beta-D-glucosidic linkages
 CC in cellobiose and cellobiose, releasing cellobiose from the non-
 CC reducing ends of the chain.
 CC -1- SIMILARITY: Contains 3 fibronectin type III domains.
 CC -1- SIMILARITY: Contains 1 bacterial-type cellulose-binding (CBD)
 CC domain.
 CC -1- SIMILARITY: BELONGS TO CELLULOSE FAMILY L (FAMILY 48 OF GLYCOSYL
 CC HYDROLASES).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.1sb-sib.ch/announce/
 CC or send an email to license@sb-sib.ch).
 CC -----
 DR EMBL: L38827; AAB00822.1; --
 DR PIR: S59077; S59077.
 DR HSSP: P07986; 1EXG.
 DR InterPro: IPR001919; Bac_cellose-bind.
 DR InterPro: IPR003961; FN III.
 DR InterPro: IPR000556; Glyco_hydro_48.
 DR Pfam: PF00553; CBM_2; 1.
 DR Pfam: PF00041; fn3_3.
 DR Pfam: PF02011; Glyco_hydro_48; 1.
 DR PRINTS: PR00844; GHYDRLASE48.
 DR ProDom: PD011903; Glyco_hydro_48; 1.
 DR SMART: SM00637; CBD_II; 1.
 DR SMART: SM00060; FN3_3.
 DR PROSITE: PS00561; CBD_BACTERIAL; 1.
 KM Cellulose degradation; Hydrolase; Glycosidase; Repeat; Signal.
 FT SIGNAL 1 33
 FT PROPEP 34 53
 FT CHAIN 54 1090 EXOGLUCANASE B.
 FT DOMAIN 54 699 CATALYTIC (BY SIMILARITY).
 FT DOMAIN 700 785 FIBRONECTIN TYPE-III 1.
 FT DOMAIN 794 884 FIBRONECTIN TYPE-III 2.
 FT DOMAIN 891 978 FIBRONECTIN TYPE-III 3.
 FT DOMAIN 989 1090 CELLULOSE-BINDING (BY SIMILARITY).
 FT ACT SITE 513 513 NUCLEOPHILE (BY SIMILARITY).
 FT DISULFID 990 1089 BY SIMILARITY.
 SQ SEQUENCE 1090 AA; 114829 MM; 046BBD956F2F399 CRC64;
 Query Match 3.38; Score 112; DB 1; Length 1090;
 Best Local Similarity 22.24; Pred. No. 5.9;
 Matches 130; Conservative 55; Mismatches 219; Indels 182; Gaps 29;
 QY 39 TGRQVSPDNFTAA-----QDLAQSLD-ANTVTFFPNISMPEFRWAK 82

DB 526 TGR-----PTMNAAPATGNGPLTVEVTSIGQVGVAAADTARALLPTAAAGSGPTASDKXK 581
 QY 83 GKIDLDSDSIRGWFYKTLDPAGATSGARAVGYSK-----IPDG----- 120
 DB 582 ALLD-----ALWANNQDPLGVS-AVETRGDKRPDDTVANODGIYISGWTGTMPND 634
 QY 121 LKPSVD-ARIRINRBCPVYTVDSVPLDGMKSLISFPMFR---TAYVAANVEN 175
 DB 635 VKPGVSFLDIFSFKKD-PNMSKVQTFIDGG-----ABQFRHREMAQTAAVAGA-- 684
 QY 176 KEMSLDVVDLLEMTNLDMRYVDSBQWIFNDTYYVIRVLRYVDPDPPFGYV 235
 DB 685 -----LADYARLPDG-----TTTDPIT-----APT--VPTGIQAGV 714
 QY 236 RYVSDRYLYKAITCEANMPL-VDQFWIGQYVALTPTSLPDYDSEAVALTHTLF--- 291
 DB 715 VTSTATISMTASTDTRVYGVYRGATKVGATATTSPTDGLTASTAVATVAFDA 774
 QY 292 ---ARPSAALAFVWAGLPQGTAPAGTPAMQASGGYLTFRANGTTPPGSVSYLP 348
 DB 775 GNVSARS--AALTVTKATPSDTPASPVAITSSSTANSVTIGMSASTDNAG----- 825
 QY 349 BGFALERYDPNDG---SWTPASAGDTVPQVAVDVRVYTNPNPAGGSGAPPTVYRVP 404
 DB 826 -DSGLAGYVYRGATRVATTLATPTDGLTASTAVETVYRARDVAGNVASSTAVSV-- 882
 QY 405 SNAYTVTVRNTLSTRPSRRLPLMPPADGQTVANNPKIEGSLNKT-----LQCY 458
 DB 883 -----TKSDTTPDT---TAPSVAGLAAMVTETSVALTNNASTDVGSGSLKGY 929
 QY 459 LVHSKRNRVFPQTPASSFGAVSFNNPG-----YERTDLPYTGIRSPQDNSTAYA 512
 DB 930 DYVR-----GATRVGSTTASYYDTGLTAATPAQYVATADNAG-----NVSA-- 973
 QY 513 HFRSLHSGSIYKTYO-----GMEG-----VTNVNT 539
 DB 974 -----SALSVTKTPQTCGSCGVANMASNMNSGPTASVRIINTGT 1014
 RESULT 11
 PIP_LACTIC STANDARD; PRT; 1902 AA.
 ID PIP_LACTIC
 AC P16271;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE P1-type proteinase precursor (BC 3.4.21.-) (Wall-associated serine
 DE proteinase).
 GN PRTP.
 OS Lactococcus lactis (subsp. cremoris) (Streptococcus cremoris).
 OG Plasmid pMW05.
 OC Bacteria; Firmicutes; Lactobacillales; Streptococcaceae; Lactococcus.
 OX NCBI_TaxId=1359;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Wg2;
 RX MEDLINE=88149035; Pubmed=3278687;
 RA Kok J., Leenhouts K.J., Haandrikman A.J., Ledebor A.M., Venema G.;
 RT "Nucleotide sequence of the cell wall proteinase gene of
 RT Streptococcus cremoris Wg2".
 RL Appl. Environ. Microbiol. 54:231-238(1988).
 CC -1- FUNCTION: PROTEASE WHICH BREAKS DOWN MILK PROTEINS DURING THE
 CC GROWTH OF THE BACTERIA ON MILK.
 CC -1- CATALYTIC ACTIVITY: Endopeptidase activity with very broad
 CC specificity, although some substrate preference have been noted,
 CC e.g. large hydrophobic residues in the p1 and p4 positions, and
 CC pro in the p2 position. Best known for its action on caseins,
 CC although it has been shown to hydrolyze hemoglobin and oxidized
 CC insulin B-chain.
 CC -1- SUBCELLULAR LOCATION: Attached to the cell wall peptidoglycan by
 CC an amide bond (Potential).
 CC -1- SIMILARITY: Belongs to peptidase family 58.

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-81b.ch/announce/
CC or send an email to license@1sb-81b.ch).
CC -----
DR EMBL; M24767; AAA17677.1; -.
DR HSSP; P00782; 1S01.
DR MEROPS; S08.019; -.
DR InterPro; IPR001899; Gram_pos_anchor.
DR InterPro; IPR006192; LPXTG.
DR InterPro; IPR003137; PA.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00746; Gram_pos_anchor; 1.
DR Pfam; PF02225; PA; 1.
DR Pfam; PF00082; Peptidase_S8; 1.
DR PRINTS; PR00723; SUBTILISTIN.
DR TIGRPFAMs; TIGR01167; LPXTG_anchor; 1.
DR PROSITE; PS50847; GRAM_POS_ANCHORING; 1.
DR PROSITE; PS00136; SUBTILASE_ASP; 1.
DR PROSITE; PS00137; SUBTILASE_SIS; 1.
DR PROSITE; PS00138; SUBTILASE_SER; 1.
KM Hydrolyase; Serine protease; Cell wall; Peptidoglycan-anchor; Zymogen;
KW Signal; Plaemid.
FT SIGNAL 1 33
FT PROPEP 34 187
FT CHAIN 188 1870
FT PROPEP 1871 1902
FT ACT_SITE 217 217
FT ACT_SITE 281 281
FT ACT_SITE 620 620
FT SITE 1867 1871
FT MOD RES 1870 1870
FT SEQUENCE 1902 AA; 199910 MW; 2901C7F1982ESD08 CRC64;

Query Match 3.3%; Score 111; DB 1; Length 1902;
Best Local Similarity 21.5%; Pred. No. 16;
Matches 123; Conservative 63; Mismatches 210; Indels 176; Gaps 31;

QY 38 RTGRQVSPDNFTAAQ---DLASLDANT---VTFPA-----NISSPERRNAMK 83
DB 776 KAGNITIVPAKTAQIEFTLSLPSFQOQFVEGFLNFKSGDSRLMLPYMGFGDWDNG 835
QY 84 KIDLDSDSIGWFKYLPAGATESARAVGEYSKIP-----118
DB 836 KIDVSLNGITY---SPAG-----GNFGIVPLTNKKTGYGYGAWTADGQOTV 882
QY 119 -DGLVKEVDAREINEECPPVTVDVSPLDGRQWLSIFSPMFRTAYVAVANENKE 177
DB 883 DDQAIARSSDK--NALYN-----DISM-----KYLLENISVQ 914
QY 178 MS-LDVNDLLEWLN---NLADKRYVDSEGINFTN---DTTYVRIKRVLPYDVPDP 230
DB 915 VDILDGCKNVTLLSSSTNLKTYNAHSQYIYVNPAMWGITY---DQ 961
QY 231 TEGLVRTVSPDYRLTYK-----AITCEANMPTL-----VDQCFWIGG 266
DB 962 RDGNIKTRADGQSYTRISGVEGDKROVPVPRKLSKAPTVNAHLSATENG---KT 1018
QY 267 QVALTFTSLPQYDVSEAYVALHTL--TFARPSAALAFVMAGLPQGGTAAPACTP--AMEQAS 324
DB 1019 QYVYLTAAB--KODLSG.DATKSVKTAINEVTNL.DATFDAGTADGYTKLETPLSDRQAQ 1076
QY 325 SCG-----YLTWRNGTTFPAGSVSYVLPBGFALERVPNGSGWTFDASADTYTR 376
DB 1077 ALGNGDNSAELVYLDNASNATDODASVQKPGSTFDL---IVNGGGIPDKISS--TTGY 1131
QY 377 QVADEVVVTNNPAGGSAFTFVRVPPS--NAYTNVTR---NTLLETPSSRRRLPLP 430
DB 1132 EA-----NTGGGCTY--TFSGTTPRAVDGTYTNAGKGDINTTYDAATNSFTASMP 1181

```

```

QY 431 MPPADP-GQTVANNPKIEGSLKRTACTLVHSMKRPVQOLTPASFGAVSFNNPGER 489
DB 1182 VTNADVAQVLDVADKAKHTLTKR-----FDRKVR-----LT-APFTDLKFNNGSDQT 1229
QY 490 TRDLDPVYGI--RDSFDQNNSTAVAFHRSLSH 519
DB 1230 SEATIKVTVGSADTKTVNVGDTVPAALDAQHH 1261

RESULT 12
GUNV_BRMCA STANDARD; PRT; 505 AA.
AC Q47096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Endoglucanase V precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase V)
DE (Cellulase V).
GN CELV.
OS Erythra carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.
OX NCBI_TaxID=554;
RX MEDLINE=94067016; PubMed=8246888;
RA Cooper V.J.C., Salmon G.P.C.;
RT "Molecular analysis of the major cellulase (Celv) of Erythra carotovora: evidence for an evolutionary 'mix-and-match' of enzyme domains."
RL Mol. Genet. 241:341-350(1993).
CC -1- FUNCTION: Endoglucanase with some exoglucanase activity.
CC -1- CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic linkages in cellulose, lichenin and cereal beta-D-glucans.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- MISCELLANEOUS: Has a pH optimum of about 7.0 and a temperature optimum about 42 degrees Celsius.
CC -1- SIMILARITY: BELONGS TO CELLULASE FAMILY A (FAMILY 5 OF GLYCOSYL HYDROLASES).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1sb-81b.ch/announce/
CC or send an email to license@1sb-81b.ch).
CC -----
DR EMBL; X76000; CA53592.1; -.
DR PIR; S39962; S39962.
DR HSSP; O85465; 1A3H.
DR InterPro; IPR001956; Glyco_hydro_5.
DR InterPro; IPR001547; CBD_3.
DR Pfam; PF00942; CBM_3; 1.
DR Pfam; PF00150; cellulase; 1.
DR PRODOM; PD001947; CBD_3; 1.
DR PROSITE; PS00659; GLYCOSYL_HYDROL_F5; 1.
KM Cellulose degradation; Hydrolyase; glycosidase; signal.
FT SIGNAL 1 31
FT CHAIN 32 505
FT DOMAIN 32 334
FT DOMAIN 335 352
FT DOMAIN 353 505
FT ACT_SITE 168 168
FT ACT_SITE 256 256
FT SEQUENCE 505 AA; 54900 MW; DBEA337BB4D2623 CRC64;

Query Match 3.3%; Score 110.5; DB 1; Length 505;
Best Local Similarity 19.0%; Pred. No. 2.5;
Matches 87; Conservative 69; Mismatches 177; Indels 125; Gaps 24;

```

```

QY 153 MSLSFSPMPTA--YVAVANVENKESLDVNDLLEMLNLA-----DMRYVDSQW 205
DB 85 MGINFRRVAMTAAGCISNPLANK-----VKRAVAAAGLVYIIIDWILSDNDEN 138
QY 206 INFNTDITVYVIRLV-----RPTVDVDPTEGLVR--TVSDYRLTYKAITCEANMP-- 255
DB 139 IYKQAKTFPAMAGLYSSPVIYIENBPQGVGMNQIPYALV-TDTRISKDDN 197
QY 256 -TLVDQGMIGGOVALTPTSLPOYDVSSEAYALH-----TLTFAPSSAALAF 302
DB 198 LIIVGTWISQDIHPLAADNQLP--DPNTWYALHFAVAGTHGFLARDIDVAGRGCAIFVS 255
QY 303 VWAGLPQGGTAPAGTPAMQASGGL-----TW-----RNGTTPAGSGSYLVLPFGPALE 354
DB 256 EM-----GTSDA-----SGNGGPELPESQTMIDFLNRRGVSWNWSLTDKSEASALA 303
QY 355 RYDPNDGSWTD--PASAGDTYFRQVAVDEVVVVNNPAGGSGAPT--TYRVPPSNAYTNT 411
DB 304 POKASGGMTEONLSTSGKFR-----EQIRAGANLGGGDTPTTPTPEPTPGNGTGD 356
QY 412 V--FRNTLLETSPSSRLLEPMPADFGQTVANNPKIEQSLK----- 452
DB 357 VLYQYRN--VDNNPDDAIRMAVNIKNTGST-----PIQLSDLYRVYFHDGKPGANLF 409
QY 453 ---ETLGCYLVSXKRNPFQULTPASSPGAVSFNN-----PGYERTDLP-----DVT 497
DB 410 VDMANVGNPNITVSTGTPTASTDKANRYLVTFSSGAGSLQPGAL-TGSHQVRIHAGWS 468
QY 498 GIRDSPDQNSTAVAHFRSLSHSCSVTKTYQGMEGVT 535
DB 469 NVNETNDYSYGANVT-----SYANMDKIT 492

```

```

RESULT 13
ID FLGE CAUCR STANDARD; PRT; 591 AA.
AC P35806;
DT 01-JUN-1994 (Rel. 29, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Flagellar hook protein flgE.
GN FLGE OR FLAK OR CC0902.
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Bacteriota; Caulobacter.
OX NCBI_TaxID=155892;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RA Mullin D.A., Mullin A.H., Newton A.;
RT "Organization and expression of Caulobacter crescentus genes needed
RT for assembly and function of the flagellar hook."
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RA MEDLINE=21173698; PubMed=1259647;
RX Nierman W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA DeRocha I., Nelson W.C., Newton A., Stephens C., Padake N.D., Ely B.,
RA Kolonay J.F., Smit U., Craven M.B., Kouri H., Shetty J., Berry K.,
RA Salberg S.L., Venter J.C., Shapiro L., Fraser C.M.;
RA "Complete genome sequence of Caulobacter crescentus."
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
RN (3)
RP SEQUENCE OF 1-52 FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=86089096; PubMed=4078896;
RA Ohta N., Chen L.S., Swanson E., Newton A.;
RT "Transcriptional regulation of a periodically controlled flagellar
RT gene operon in Caulobacter crescentus."

```

```

RL J. Mol. Biol. 166:107-115(1985).
CC -1- SIMILARITY: BELONGS TO THE FLAGELLA BASAL BODY ROD PROTEINS
CC FAMILY.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announce/
CC or send an email to license@ebi.ac.uk.)
CC -----
DR EMBL, AF072135, AAC33329.1; -.
DR EMBL, AB005767, AAX22866.1; -.
DR EMBL, M28551, AAA23042.1; -.
DR PIR, B87361, B87361.
DR PIR, I40664, I40664.
DR TIGR, CC0902, -.
DR InterPro, IPR001444, Flag_bb_rod.
DR Pfam, PF00460, flag_bb_rod, 1.
DR PROSITE, PS00588, FLAGELLA_BB_ROD, 1.
DR KMW, Complete proteome.
SQ SEQUENCE 591 AA; 61207 MW; PD4BEA319479A161 CRC64;

Query Match 3.2% Score 109; DB 1; Length 591;
Beet Local Similarity 19.6% Pred. No. 4; Mismatches 236; Indels 220; Gaps 31;
Matches 130; Conservative 76; Mismatches 236; Indels 220; Gaps 31;

14 RRGTRNVSVANTVYVNGRRQRRTGRQVSPDNPFAADLQSLDN----- 63
37 KRSTSNF-----STLVTSNKNQYTSAGGVKQTHQF-ISQGLTSTSTNSDISGAGF 91
64 --TTPPANISMPFRNMAK-GKIDLS-----DSIMYRK-VY-----DPAGA 104
92 FYTTEKPEMLTA-TDTRSFTRAGSPQDNLGYLRNDAGLYLQGMADPVSGLTPDPSDL 150
105 TESARA-VGEYSKIPDGLVKPSVDABIREINBECPPVTDVSVPLDGRQWSLSTSPMF 163
151 MQLASINVGSGVGAETKTRGVANANLR-----SEQPVAAASVYK----- 190
164 RTAVVAVANVENKESLDVNDLLEMLNLAIDMRYVDSQWINFNTDITVYVIRVLRP 223
191 ---VGTAGSPBK-----TNVDSA-----TNS-----H 210
224 TYDVPDPTEGLVRYSDYRLTYKAITCEANPFLVDQGMIGQYALTPSLPOYDSEBA 283
211 NYDVVYSSGTIANPVS-----GNNBYLVD--IKENGVIATGKVAVD----- 250
264 YALHTLTFARPSAALAFVWAGLPQGGTAPAGTPAMQASGGLTWVRHN--GTFPAG 341
251 -----AATNELVSSITDYKASPV-----TGSMTTRINAAGTTVNLA 288
342 SVSYVLPEG-----FALERYDPNDGSWTDFASAGDTYFRQVAVDEVVVVNNPAGGSGA 395
289 DLGIVNAGADABRVVAGKLYDPSITWMSMDYAKONSK-----GVX 328
336 PFTYRVPPSNAY--TNTVFRNTLLETSPSSRLLEPMPADFGQTVANNPKIEQSLK 453
339 PPFYQIPLPSOSKGGQRTVTLTSMKGPSPNQYAEIRAKFGDLANN--GNGQISGTIIE 385
454 TLGCVLVSXKRNPFQULTPASSPG-----AVSFNNPGYERTDLP-----DY 496
366 ----FTTQOKLKN-----TGSLEFPTTSPTALTITISGVIAPVTTPPAVQPTPTPWADA 435
497 TGIRSPDQNSTAVAHFRSLSHSCSVTKTYQGMEGVTNNTVTPGQFAHAGLKNRBITL 556
436 LGIDQBEVQIDLASAAGGLQVNSQSV-----QSVNTGTATAGNLINIEIDGGYVS 488
557 CLADDLATRLTGVYATDNPFAAVSAPA-ANMLSSV-----LKSEATSSITKSVEGTAV 609
489 ALFDDGVVTRRLA-----QVALTFSPNPLKGVNGNAVYVTSBGSTYSLKAPSOQGA 540
610 GA 611

```

Db 541 GA 542

RESULT 14
ID GUNN_ERWCA STANDARD; PRT; 444 AA.

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Endoglucanase N precursor (EC 3.2.1.4) (Endo-1,4-beta-glucanase N) (Cellulase N).

GN CELN.
OS Erwinia carotovora.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Pectobacterium.

OX NCBI_TaxID=554;

RP SEQUENCE FROM N.A.

RC SMPAIN-Acrosepica FCBR C18;
RX MEDLINE=9629944; PubMed=9636315;
RA Olsen O., Thomsen K.K., Weber J., Dues J.O., Svendsen I., Wegener C.,
RA von Wettstein D.;

RT "Transplanting two unique beta-glucanase catalytic activities into
RT one multi-enzyme, which forms glucose.";

RL Biotechnology 14:71-76(1996).

CC -1 CATALYTIC ACTIVITY: Endohydrolysis of 1,4-beta-D-glucosidic
CC linkages in cellulose, lichenin and cereal beta-D-glucans.

CC -1 SUBCELLULAR LOCATION: Secreted.

CC -1 SIMILARITY: BELONGS TO CELLULOSE FAMILY A (FAMILY 5 OF GLYCOSYL
CC HYDROLASES).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: L39788; AAC37033.1; -

DR HSSP: O85465; IASB.

DR InterPro: IPR001956; Glyco_hydro_5.

DR InterPro: IPR001547; Glyco_hydro_5.

DR Pfam: PF00942; CBM_3; 1.

DR Pfam: PF00150; cellulase; 1.

DR ProDom: PD001947; CBD 3; 1.

DR PROSITE: PS00659; GLYCOSYL_HYDROL_F5; 1.

KM Cellulose degradation; Hydrolase; Glycosidase; Signal.

FT SIGNAL 1 31 POTENTIAL.

FT CHAIN 32 444 ENDOLUCANASE N.

FT ACT_SITE 168 168 PROTON DONOR (BY SIMILARITY).

FT ACT_SITE 256 256 NUCLEOPHILE (BY SIMILARITY).

SO SEQUENCE 444 AA; 48300 MW; PA7E4179004CB843 CRC64;

Query Match 3.2%; Score 108.5; DB 1; Length 444;

Best Local Similarity 20.9%; Pred. No. 2.9;

Matches 70; Conservative 48; Mismatches 138; Indels 79; Gaps 19;

QY 153 WSLISFSPMPRTA--YVAANVENKESLDVVDLEWLNLA-----DKRYVDSQW 205

Db 85 WGVNFEVAMTADGDIISNPLANK-----VEAVAAOSLGVIITIDHIIISDNDPN 138

QY 206 INFNTDITYVIRIVL-----RPTVDVDPTEGLVR--TVSDYRLTYKAITCEANMP-- 255

Db 139 IYKAQATPFAMMGVSSPNVIYBIANEPNGVTWNGQIRPALAEV-TTIRSKDPDN 197

QY 256 -TLVDQGFNIGGYALPTSLPQYDVSEAYALH-----TLTPAPSSAALAF 302

Db 198 LIITGTGTWSDIHDAADNDLP--DPNTLYALHFYAGTHGQFLDRIDVAGSRGALFVS 255

QY 303 VMAGLPGQGTAPAGTAPWAGASGGYL---TW-----RHNGTTPAGSVSVLPEGFAL 354

Db 256 EM-----GTSDA-----SGNGCPFLPESQTRIDFLANRGVSMVMTTKSEASALIA 303

QY 355 RYDNDGSGWTD--FASAGDTVTFRQVAVDEVVYTNNPAGSGSAPFTVRVP-----PSNA 407

Db 304 FGASKSGMTKEQNISTSGKFRV-----EQRAGAMIGGGDFTT-TPTTPTPTPNONG 355

QY 408 YTNV---FNTLTLETPSSRLLEPMPADPQGT 439

Db 356 TTGDVVLQYRN--VDNNPSDDAIRMAVNIKNTGST 388

RESULT 15

ID VP4_ROT3 STANDARD; PRT; 775 AA.

AC P39033;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Outer capsid protein VP4 (Hemagglutinin) (Outer layer protein VP4)

DE [Contains: Outer capsid proteins VP5 and VP8].

GN S4.

OS Human rotavirus (serotype G3 / strain AU-1).

OC Viruses; dsRNA Viruses; Reoviridae; Rotavirus.

OX NCBI_TaxID=39013;

RP SEQUENCE FROM N.A.

RX MEDLINE=92356070; PubMed=1322955;

RA Iseigawa Y., Nakagomi O., Nakagomi T., Ueda S.;

RT "A VP4 sequence highly conserved in human rotavirus strain AU-1 and
RT feline rotavirus strain FRV-1.";

RL J. Gen. Virol. 73:1939-1946(1992).

CC -1 SUBCELLULAR LOCATION: Outer capsid.

CC -1 PFM: VP8 IS ONE OF TWO TRYPSIN CLEAVAGE PRODUCTS OF VP4; THE
CC OTHER PRODUCT IS VP5.

CC -1 SIMILARITY: BELONGS TO THE ROTAVIRUS VP4 PROTEIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC EMBL: D10970; BA01747.1; -

DR PIR: J01638; J01638.

DR InterPro: IPR00416; Cap_VP4.

DR Pfam: PF00426; VP4; 1.

KM Coat protein; Glycoprotein.

FT CHAIN 1 775 OUTER CAPSID PROTEIN VP4.

FT CHAIN 1 775 OUTER CAPSID PROTEIN VP8.

FT CARBOHYD 17 17 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 32 32 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 97 97 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 132 132 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 183 183 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 198 198 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 237 237 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 567 567 N-LINKED (GLCNAC...) (POTENTIAL).

FT CARBOHYD 613 613 N-LINKED (GLCNAC...) (POTENTIAL).

SO SEQUENCE 775 AA; 87092 MW; 4148D0C91A89B18A CRC64;

Query Match 3.2%; Score 108; DB 1; Length 775;

Best Local Similarity 20.9%; Pred. No. 7;

Matches 67; Conservative 51; Mismatches 132; Indels 70; Gaps 14;

QY 324 SSGGYLTWRANGTTPFAGSVSYLPSGFALERDP-NDGS-----NTDFASAGDTVTR 376

Db 319 SVNGVNDENPENGFT-----LPTDFALSRFEVINKENSVYVYVDDSDQAFRMYVVR 369

QY 377 QVA--VDEVV-----VTNNPAGGGSAPFTT-VRVPSSNAVNTVYFRNTLTETRP 422

```

Db      370  SLAANTLVYCSGGSYSPALPVGNHFPVMSGAVTITSAGVTITSTQYTDYVSLNSL----- 424
Qy      423  SSRLELPMPPADFCQTVANNPKIRQSLKETLGCYLVHKKRNPFQLTTPASSFGAVSF 482
Db      425  -RFRRLAVSEPSF-----SISRTMSGIYGLPAVNPNNNAEYELIAGRFLISLVL 475
Qy      483  NNPGERTRDLPDYTGIRDSFPQNMSTAVAHFRSLSHSCSIYTKTYQMEGVNTVNTPPG 542
Db      476  TNDYQ--TPIANSVTVRODLERQJELKEEFNSLSOEIAVS-----QLIDLATLPID 526
Qy      543  QFAHAGLKN--EBILCLADLATRLTGVPATDNFAAASAFANMLSVLKSEATSSI 600
Db      527  MFSMFSGIKSTYEAVKSKMTTNMKRF-----KTSLSLANHISDLTSM-----SEAASSV 575
Qy      601  ---IKSVGETAVGAAGSGL 616
Db      576  RLTSVRSYGVTVLPPRAVSL 595

```

Search completed: January 15, 2004, 16:50:07
 Job time : 14.8065 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 16:06:47 ; Search time 25.869 Seconds
(without alignments)
6454.060 Million cell updates/sec

Title: US-09-991-262-50

Perfect score: 3374

Sequence: 1 MGDAGVASQRPNRRGTRNV.....GKTAARVARRARRAARAN 647

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL.23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriaph:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3374	100.0	647	12	Q82462
2	2205	65.4	12	12	Q82462
3	435.5	12.9	12	12	Q90063
4	407.5	12.1	12	12	Q9YK87
5	386.5	11.5	12	12	Q8OYMS
6	140.5	4.2	12	12	Q9YR82
7	134	4.0	12	12	Q8YR84
8	132	3.9	12	12	Q9L950
9	131.5	3.9	12	12	Q9Q887
10	131	3.9	12	12	Q56854
11	130	3.9	12	12	Q8EV71
12	127	3.8	12	12	Q9YR11
13	127	3.8	12	12	Q95810
14	127	3.8	12	12	Q95TGO
15	127	3.8	12	12	Q8MKW7
16	127	3.8	12	12	Q9NBA1

17	127	3.8	2019	5	Q8MKM8	Q8MKM8 drosophila
18	125.5	3.7	717	2	Q9KRB3	Q9KRB3 bacillus cl
19	125.5	3.7	1269	17	Q97Z06	Q97Z06 sulfolobus
20	125	3.7	2314	16	Q69822	Q69822 streptomyces
21	124.5	3.7	2283	9	Q8VQ99	Q8VQ99 staphylococcus
22	123.5	3.7	1204	17	Q9UY51	Q9UY51 pyrococcus
23	123	3.6	5020	16	Q8E9W3	Q8E9W3 shewanella
24	122	3.6	1093	16	Q8EV69	Q8EV69 mycoplasma
25	122	3.6	1314	4	Q9UMZ2	Q9UMZ2 homo sapien
26	122	3.6	1582	16	Q8I9A5	Q8I9A5 listeria mo
27	121.5	3.6	699	2	Q48494	Q48494 kurtzia zop
28	121	3.6	958	17	Q8TL14	Q8TL14 methanobac
29	120.5	3.6	769	12	Q9PZT4	Q9PZT4 human parvo
30	120.5	3.6	781	12	P89318	P89318 human parvo
31	120.5	3.6	979	2	Q9XAS7	Q9XAS7 streptococcus
32	120	3.6	873	16	Q9RX36	Q9RX36 deinococcus
33	120	3.6	1091	2	Q8KRP7	Q8KRP7 paenibacill
34	119.5	3.5	851	12	Q9QF86	Q9QF86 human herpe
35	119.5	3.5	889	4	Q14917	Q14917 homo sapien
36	119.5	3.5	907	12	Q6E537	Q6E537 human herpe
37	119.5	3.5	1496	5	Q9VIT9	Q9VIT9 drosophila
38	119.5	3.5	2271	16	Q99QY4	Q99QY4 staphylococ
39	119	3.5	1993	16	Q92FG5	Q92FG5 listeria in
40	119	3.5	2809	5	Q61230	Q61230 lytechinus
41	119	3.5	22152	4	Q8WX17	Q8WX17 homo sapien
42	118.5	3.5	1743	17	Q8RT17	Q8RT17 methanobac
43	118.5	3.5	2566	17	Q8TSE7	Q8TSE7 methanobac
44	118.5	3.5	3029	16	Q55582	Q55582 synchocyst
45	117.5	3.5	781	12	Q9VGP8	Q9VGP8 human parvo

ALIGNMENTS

RESULT 1
ID Q82462 PRELIMINARY; PRT; 647 AA.

AC Q82462;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Coat protein.
GN P71.
OS Helicoverpa armigera stunt virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
OC Unclassified Tetraviridae.
OC NCBI_TaxID=37206;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=black mountain;
RX MEDLINE=97201539; PubMed=9049325;
RA Hanzlik T.N., Dorian S.J., Johnson K.N., Brooks E.M., Gordon K.H.;
RT "Sequence of RNA2 of the Helicoverpa armigera stunt virus
(Tetraviridae) and bacterial expression of its genes.";
RL J. Gen. Virol. 76:799-811(1995).
DR EMBL; L37299; AAC37885.1; --
DR MEROPS; A21.001; --
DR InterPro; IPR005313; Peptidase A21.
DR Pfam; PF03566; Peptidase A21.1.
SQ SEQUENCE 647 AA; 70670 MW; 0B205B8CB53CFBD2 CRC64;

Query Match 100.0%; Score 3374; DB 12; Length 647;
Best Local Similarity 100.0%; Pred. No. 6.4e-239; Index 0; Gaps 0;
Matches 647; Conservative 0; Mismatches 0;

QY 1 MGDAGVASQRPNRRGTRNVASANTVTYVNGRGNORRGTGROYSPDNPFAAODLAQSL 60
|||||
DB 1 MGDAGVASQRPNRRGTRNVASANTVTYVNGRGNORRGTGROYSPDNPFAAODLAQSL 60
|||||
QY 61 DANTVTTPANISSMPEFRNNAKGIIDLSDSISGMVFRYLDPAQATSSARAIVGYSKIPDG 120
|||||
DB 61 DANTVTTPANISSMPEFRNNAKGIIDLSDSISGMVFRYLDPAQATSSARAIVGYSKIPDG 120
|||||

```

QY 121 LVKSVDAEIRIYNEECPPVTVDSVPLDGRQMSLSTSPFMRFRATYAAVANVENKMSL 180
DB 121 LVKSVDAEIRIYNEECPPVTVDSVPLDGRQMSLSTSPFMRFRATYAAVANVENKMSL 180
QY 181 DVVNDLIMLNNLAWRYVVDSEQWINFNDTTYVYR.VYLR.PTYDVDPREGIVRTVSD 240
DB 181 DVVNDLIMLNNLAWRYVVDSEQWINFNDTTYVYR.VYLR.PTYDVDPREGIVRTVSD 240
QY 241 YRLTYKAITCEANMPTLVDOGFWMIGQYALPTPTSLPOYDVSEAVYALHTLTPARPSAAAL 300
DB 241 YRLTYKAITCEANMPTLVDOGFWMIGQYALPTPTSLPOYDVSEAVYALHTLTPARPSAAAL 300
QY 301 AFVWAGLPQGGTAPAGTAPAMEQASSGGYLTWRHNGTTPAGSVSYLPEGFALERYDND 360
DB 301 AFVWAGLPQGGTAPAGTAPAMEQASSGGYLTWRHNGTTPAGSVSYLPEGFALERYDND 360
QY 361 GSWTDFASAGDTVTRQVAVDEVVYVNNPAGGSA.PTFVVRPBNATNTVFRNTLLBT 420
DB 361 GSWTDFASAGDTVTRQVAVDEVVYVNNPAGGSA.PTFVVRPBNATNTVFRNTLLBT 420
QY 421 RPSRRLELPMPPADPGQTAVANNPKIEOSLKEITLGCYLVHSMKMPVFOITPASSFGAV 480
DB 421 RPSRRLELPMPPADPGQTAVANNPKIEOSLKEITLGCYLVHSMKMPVFOITPASSFGAV 480
QY 481 SPNNPGYERTDLPDYTGIRDSFDQNMSTAVAHFRSLSHSGSVTKTYQMGEGVTVNTP 540
DB 481 SPNNPGYERTDLPDYTGIRDSFDQNMSTAVAHFRSLSHSGSVTKTYQMGEGVTVNTP 540
QY 541 FGQFPAAGLKNKEEILCLADDLATRLTGYYPATNDPAVAASFAANMLSSVLSKRSATSI 600
DB 541 FGQFPAAGLKNKEEILCLADDLATRLTGYYPATNDPAVAASFAANMLSSVLSKRSATSI 600
QY 601 IKSVEYAVGAAGSGLAKLPGILMSVPGKIAARVBARARRABAN 647
DB 601 IKSVEYAVGAAGSGLAKLPGILMSVPGKIAARVBARARRABAN 647

```

RESULT 2

```

Q90063 ID 090063 PRELIMINARY; PRT; 644 AA.
AC 090063:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)
DE Capsid protein.
GN CAPSID PROTEIN, CP.
OS Nudavirelia capensis omega virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
OC Omegatetravirus.
OC NCBI_TaxID=12541;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=9231097; PubMed=1519360;
RA Agrawal D.K., Johnson J.E.;
RT "Sequence and analysis of the capsid protein of Nudavirelia capensis
RT omega virus, an insect virus with T = 4 icosahedral symmetry.";
RL Virology 190:806-814(1992).
DR EMBL; S43937; AAB23198.1; -.
DR MEMOPS; A21.001; -.
DR InterPro; IPR005313; Peptidase_A21.
DR Pfam; PF03566; Peptidase_A21; 1.
SQ SEQUENCE 644 AA; 69863 MW; EA41G9AB5OF24828 CRC64;

```

```

Query Match 65.4%; Score 2205; DB 12; Length 644;
Best Local Similarity 65.8%; Pred. No. 4e-153;
Matches 427; Conservative 69; Mismatches 135; Indels 18; Gaps 6;

```

```

QY 8 SQRPHNRGTRNVRVSAANTVTYNGRRNR--RRTGRQVSPDNFTAAADLAQSDANTV 65
DB 3 SNSASGKRSRNVRVRIANTVNAVAPKQARGRARRRANININVTAAAELOGLSDANVI 62
QY 66 TTPANISSMPEFRNNAKGIIDSDISIGVYFYLDAGTSAARAVGEXSKI.PDGLVKFS 125

```

```

DB 63 TTPNVAATMPEFRSMARKGLDIDDSIGWYKYLDPAGATSAAGVGRYSKI.PDGLVKFS 122
QY 126 VDABIRIYNEECPPVTVDSVPLDGRQMSLSTSPFMRFRATYAAVANVENKMSL 185
DB 123 VDABIRIYNEECPPVTVDSVPLDGRQMSLSTSPFMRFRATYAAVANVENKMSL 182
QY 186 LIEMNLANTADRRYVVDSEQWINFNDTTYVYR.VYLR.PTYDVDPREGIVRTVSD 245
DB 183 LIEMNLANTADRRYVVDSEQWINFNDTTYVYR.VYLR.PTYDVDPREGIVRTVSD 242
QY 246 KAITCEANMPTLVDOGFWMIGQYALPTPTSLPOYDVSEAVYALHTLTPARPSAAALAFWA 305
DB 243 KAITCEANMPTLVDOGFWMIGQYALPTPTSLPOYDVSEAVYALHTLTPARPSAAALAFWA 302
QY 306 GLPQGGTAPAGTAPAMEQASSGGYLTWRHNGTTPAGSVSYLPEGFALERYDND 365
DB 303 SMPGGSAPSGDDPAMIPSTTQF-QMHRIGSDATPTGVITTYIPGTYMQYDITTNBNG 361
QY 366 PASAGDTVTRQVAVDEVVYVNNPAGGSA.PTFVVRPBNATNTVFRNTLLBT 418
DB 362 PASAGDTVTRQVAVDEVVYVNNPAGGSA.PTFVVRPBNATNTVFRNTLLBT 413
QY 419 ETRPSRRLELPMPPADPGQTAVANNPKIEOSLKEITLGCYLVHSMKMPVFOITPASSFG 478
DB 414 ETRPSRRLELPMPPADPGQTAVANNPKIEOSLKEITLGCYLVHSMKMPVFOITPASSFG 473
QY 479 AVSFNPNPGYERTDLPDYTGIRDSFDQNMSTAVAHFRSLSHSGSVTKTYQMGEGVTVN 538
DB 474 AVSFNPNPGYERTDLPDYTGIRDSFDQNMSTAVAHFRSLSHSGSVTKTYQMGEGVTVN 533
QY 539 TTPGQFPAAGLKNKEEILCLADDLATRLTGYYPATNDPAVAASFAANMLSSVLSKRSATSI 598
DB 534 TTPGQFPAAGLKNKEEILCLADDLATRLTGYYPATNDPAVAASFAANMLSSVLSKRSATSI 593
QY 599 SIISVGEYAVGAAGSGLAKLPGILMSVPGKIAARVBARARRABAN 647
DB 594 SIISVGEYAVGAAGSGLAKLPGILMSVPGKIAARVBARARRABAN 642

```

RESULT 3

```

Q9YK87 ID 09YK87 PRELIMINARY; PRT; 757 AA.
AC 09YK87:
DT 01-MAY-1999 (TREMBLrel. 10, Created)
DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Capsid protein.
GN TAV-CP.
OS Thossea asigna virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;
OC Unclassified Tetraviridae.
OC NCBI_TaxID=83810;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99350019; PubMed=10423156;
RA Pringle F.M., Gordon K.H., Hamzlik T.N., Kalmakoff J., Scotti P.D.,
RA Ward V.K.;
RT "A novel capsid expression strategy for Thossea asigna virus
RT (Tetraviridae).";
RL J. Gen. Virol. 80:1855-1863(1999).
DR EMBL; AF062037; AAC97195.1; -.
DR InterPro; IPR005313; Peptidase_A21.
DR Pfam; PF03566; Peptidase_A21; 1.
FT CHAIN 1 155 UNKNOWN.
FT CHAIN 156 695 LARGE CAPSID PROTEIN.
FT CHAIN 696 757 SMALL CAPSID PROTEIN.
SQ SEQUENCE 757 AA; 82413 MW; 17432C707D2C427C CRC64;

```

```

Query Match 12.9%; Score 435.5; DB 12; Length 757;
Best Local Similarity 25.9%; Pred. No. 3.8e-23;
Matches 170; Conservative 105; Mismatches 263; Indels 119; Gaps 27;

```

```

QY 26 TYTVNAGRNRNR--TGRQVSPDNFTAAADLAQSDANTVTTPANISSMPEFRNNAK 82

```



```

Db      174 TKAPGKTRRRRRNGKRI GNPQPSISA-----GAPMSKIRPLKCTLA 219
Qy      83 GKIDLSDSIGMYFKYLDPAAGATESARAVEYSKIPDGLVKFSYDAIREIYNEBCEPVVT 142
Db      220 MTEPIPSISGIMLETNLDPCGEYKSTL---DYGVPRGAI PMSVCGFRFTPI RHGAN 276
Qy      143 DVSVPIDGROWMSISFEPMPRTAYAVANVENKMSLDVNDLIEMLAN---LADWRY 198
Db      277 PMLAPIDGSMWTGII RTNLMRTPEFLIADMMNAEISDESIDDEVARFNNYVGLDAD--- 333
Qy      199 VVDSEQMINFTNTTYYVIRVLRPTVDOPTE-GLVRTVSDRYLTAKITCEANMPTL 257
Db      334 -ASHPNMVE-TSQAGLYWSIQRTALIGVPPSDTGSPIYTDPRITGSGFTVSSNTPSL 391
Qy      258 VDOGFWIGGOYALTPTS-----LPQYDSEAYALHTLTTPARPSAALAFVWAGLPQGGT 312
Db      392 INGMVYVVAQF--NPNTENKEVIPHSESGETPLGMSRLVSTASTVALITYIQGV--CGA 447
Qy      313 APAGTPAMEQASGGYITWRHNGTTFPAGSV--SYVLEGPALERYDPNDGSMWTFASAG 370
Db      448 QNORTV-----TLNRSQSVINSSVLPQ--VTQGYAFNNG---DTYAVG 485
Qy      371 DVTTFPQVAVDEVVNTNPNAGGSAPTFTVRVPS--NAYTN-----TVFRN----- 415
Db      486 DTLRFQ-----LTERDSGG-----YTVRLQRSADGTTFTNIGPLSTTGQNGFNPD 532
Qy      416 ----TLETFRSSRLLE-LPMPADFGQTVANNPKIEQSLKETIGCYLVHSMKRNVPFQ 470
Db      533 SGIGVILLEGVLAGRVNMLTLPPTQSDLMQOTPKTCVPOIKE--GGFYVRQDIWQVYN 590
Qy      471 LTPASSFGAVSFNNPGYERTD-LPDYTG-IRDSFDONMSTAVAHFRSLSHSGSVTKTY 528
Db      591 MTPASRYAPRFYSNAI--TLDAINSAVGTIRDTADSNYFALCHMTSLPLACAPFIKAS 648
Qy      529 QGREGVNTVTPGQFPAHGLKNEELICLADDLATLTGVPATONFAANVASFANML 588
Db      649 MREBAAPGRNSPMPGFTSTPPDEVALTTSRYVMDEIEPPAMPPLAYNGMGLMSKVLSTIV 708
Qy      589 -SSYLKSEATSIKSVGETAVGA---AQSLAKLPGLMSVPGKIARVBARARR 642
Db      709 RRIPLKLRKAS---NNAETVAGCYEDNAQAVT-----KVYETVAGRGAMR 751

```

RESULT 4

O8QVMS

ID O8QVMS PRELIMINARY; PRT; 756 AA.

AC O8QVMS; (TREMBLrel. 21, Created)

DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)

DE 01-OCT-2002 (TREMBLrel. 22, Last annotation update)

DS Capsid protein.

OS Euprosterne elaeasa virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage.

OX NCBI_Taxid=186165;

RN [1]

RP SEQUENCE FROM N.A.

RA Zeddam J.-L.A., Pringle F.M., Gordon K.H., Ward V.K., Luke B.T.,

RB Gorbaleya A.E., Hanzlik T.N.;

RT "Genome organization of Euprosterne elaeasa virus defines it as a

RL member of a new group of insect RNA viruses."

RU Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF461742; AAL78971.1; "

DR InterPro; IPR005313; Peptidase A21.

DR Pfam; PF03566; Peptidase_A21; I.

FT CHAIN 1 150 SMALL PUTATIVE PEPTIDE.

FT CHAIN 151 695 LARGE CAPSID PROTEIN.

FT CHAIN 696 756 SMALL CAPSID PROTEIN.

SQ SEQUENCE 756 AA; 82698 MW; 1061486A17DB20BC CRC64;

Query Match 12.1%; Score 407.5; DB 12; Length 756;
 Best Local Similarity 24.6%; Pred. No. 4.3e-21;
 Matches 154; Conservative 100; Mismatches 302; Indels 69; Gaps 20;

```

Qy      32 RRRRRRTGQVPEPPDFTAAADLAQSLDANTVTTPANISSMPEFPNNAKIKIDLSDS 91
Db      181 RRRRRRTVGTGNPN-----OPFA-----VTGAAALTRVPLALGTLAMESEIPSS 227
Qy      92 IGMVFKYLDPAAGATESARAVEYSKIPDGLVKFSYDAIREIYNEBCEPVVTDSVPLDR 151
Db      228 VGMLESTNLDPCGEYKSTL---DYGVPRGAI PMSVCGFRFTPI RHGANPVAFLDGT 284
Qy      152 QMSLISFSPMPRTAYAVANVENKMSLDVNDLIEMLNLDWRYVVDSEQMINFTND 211
Db      285 MMTLGIIIRTNLMRTPEFLIADMMNAEISDSGLBVARSFNNYGEIELASHPMVETLQE 344
Qy      212 TTYVIRVLRPTVDOPTE-GLVRTVSDRYLTAKITCEANMPTLVDOGFWIGGOYAL 270
Db      345 GLYW-SIQRTALIGVPPSDTGSPIYTDPRITGSGFTVSSNTPSLINGMVVAQF-- 401
Qy      271 TPTS-----LPQYDSEAYALHTLTTPARPSAALAFVWAGLPQGGTAPAGTAMEQASS 325
Db      402 NPNTENKEVIPHSESGETPLGMSRLVSTASTVALITYIQGV--GGAQNTFTVI--ANR 456
Qy      326 GGYLTRHNGTTPAGSVSYVLEGPALERYDPNDG---SWTFASAGDVTFRQVAVDR 382
Db      457 GPAVI---NSGVLPAYTQGRFLNG---DTYAVGDSLRVQFTERDGGTYV-YLQSTDG 509
Qy      383 VVVTNNPAGGGSAPTFTVRVPSNAYTNT---VFRNTLLETFRSSRLLELPMPADFGQ 438
Db      510 TTFETN-----IGPPTTAAAPGDTGTAAGVGFILBGDVLAGVNM---LTLPPFQSD 559
Qy      439 TVANNPKIEQSLKETIGCYLVHSMKRNVPFQLTTPASSPFAVSPNNGYERTDLPYTG 498
Db      560 LMQOTPKTCVPOIKE--GGFYVRQDIWQVYNMTPASRYAPRFYS- GAITLDSLNSVG 616
Qy      499 -IRDSFDONMSTAVAHFRSLSHSGSVTKTYQGMGVTNNTFGQFAHGLKNEELIC 557
Db      617 TIRDTADSNYFALCHMTSIFLACAPFIKAMFBAVPGNSWGFLPSTTPEDVRLT 676
Qy      558 LADDLATRLTGVPATONFAANVASFANMLSSVLKSEATSIKSVGETAVGAQSLA 617
Db      677 ISRTVMDIEPPAMPPLAYNGMGLMSKIKINIIRKVPKYLRTASVADVANCVEDAAGAT 736
Qy      618 KLPGLMSVPGKIARVBARARR 642
Db      737 -----RYESVGRRRGAMR 750

```

RESULT 5

O9YRB2

ID O9YRB2 PRELIMINARY; PRT; 612 AA.

AC O9YRB2; (TREMBLrel. 10, Created)

DT 01-MAY-1999 (TREMBLrel. 10, Last sequence update)

DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)

DS Capsid protein.

OS Nudaurelia capensis beta virus.

OC Viruses; ssRNA positive-strand viruses, no DNA stage; Tetraviridae;

OX NCBI_Taxid=85652;

RN [1]

RP SEQUENCE FROM N.A.

RA MEDLINE=9263183; Pubmed=10329566;

RB Gordon K.H., Williams M.R., Hendry D.A., Hanzlik T.N.;

RT "Sequence of the genomic RNA of nudaurelia beta virus (Tetraviridae)

RL defines a novel virus genome organization."

RU Virology 258:42-53(1999).

DR InterPro; IPR005313; Peptidase A21.

DR Pfam; PF03566; Peptidase_A21; I.

SQ SEQUENCE 612 AA; 66406 MW; 0B34029D23B39085 CRC64;

DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
 DE Outer membrane-like protein.
 OS Pseudomonas putida.
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
 OC Pseudomonadaceae; Pseudomonas.
 OX NCBI_TaxID=303;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=20225830; PubMed=10762233;
 RA Bepinosa-Urgel M., Salido A., Ramos J.L.;
 RT "Genetic analysis of functions involved in adhesion of pseudomonas
 RT putida to seeds";
 RL J. Bacteriol. 182:2363-2369(2000).
 RN [2]
 RN SEQUENCE FROM N.A.
 RA Bepinosa-Urgel M., Salido A., Ramos J.L.;
 RL Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF182515; AAF69025.1; -; BD7FC3B460FE2D6B CRC64;
 SQ SEQUENCE 2147 AA; 211587 MW; BD7FC3B460FE2D6B CRC64;
 Query Match 4.0%; Score 134; DB 2; Length 2147;
 Best Local Similarity 22.0%; Pred. No. 2.7;
 Matches 141; Conservative 75; Mismatches 264; Indels 162; Gaps 33;
 QY 39 TGRQVSPD-NFTAAADLAQSIDANTVTPPANISSMPEFRNNAKKIDSDSI--GM 94
 DB 1363 TGTALPAPDLTPPAAAGNLFFSADGTLTGTGAGATVTVRS-AGGTV-LQTATVQADGT 1420
 QY 95 YFVYLDPAQATESARAVGEYSKI PDGLVPSVDAEIREINSCPVYTD-----VSFPLD 149
 DB 1421 FTVYTLSP--PDQGQILSVTLSDPRGVASGVNAVTVADV-DANAVIASNTLATATVLA 1477
 QY 150 G-----RQMSLSIFSPMFRTAYAVVANVENKMSLDVNDLLEMLNLADWRYVD 201
 DB 1478 AVSATKTYTDSFTLLSGFKTYTWVAAGTADPTLT--TTSVSLALANTSPTLQVKD 1536
 QY 202 -SEQWINEFTNDTYVYRIRVLRPTVDPTBEGI-----VRYSDVRLTYKA--ITCEAN 253
 DB 1537 ASGAWVTIATGNTQGLDLTLVLP-----GLQVDIGVLQAGDYRLTVGSGGIGLITE 1588
 QY 254 MPTLVDOGFWIGQGYALPTSLPOY-----DYSEAVALTHTLTFAR-PSSAA 239
 DB 1589 VSTTLTLD-----IVSTSLTQFTGTGATSGNVITDVGTDGAVD---AGPDSAAV 1634
 QY 300 LAFVWAGLPQGTAPAGTAPMEQASG-----GYLTMHNGTTFPAGSV--SYVL-- 347
 DB 1635 LQV-----LKDGSVYAGTATTVQGYGTIVIRADGSYTYTPNGSPNSVGKVDVPSYQVH 1690
 QY 348 PEGF--ALERY-----DPDGSWTFPASAGDTVTRQVAVD-EVVVTNPNAG 391
 DB 1691 PNGLSAANLYVRIDSPQATEVMSDTMYGAPATVADVANDVGTISQVTLDRVSSSSSTLG 1750
 QY 392 GGSAPFTTVVPPSNMAYNTVFPNTLLETTPSSRRLLEPMPADFGQTVANNPRIEBSL 451
 DB 1751 SLNLVPIGTR--CATTTTVAANTTADLVVNSTNL-----LSLNGTTTELK 1799
 QY 452 KETLGCYLIVHAKRNPFVOLTASSFGAVSPNNPGYERTDLDPYTGIRSDPQNNSTAV 511
 DB 1800 NPATGQV-----LVQSVPGSLVSIIGGAGY-----TF-ENGAGT 1836
 QY 512 AFRSLSHSCSIVTKTYQMEGVNTVNP--FGQFAHAGLKNBEILLADLALRFLG 568
 DB 1837 YHVRVYAGIGLSSSI-----TTSINTTTTYLTFEYVSG-----ATATG 1876
 QY 569 VVPATDNFAAVGAFANMLSVLKSAATSIISKVETAVG 610
 DB 1877 NLLADNVGSALTIVL-----SVLTANTYTTIPIQYNGSVAG 1912

ID 09Q87 PRELIMINARY; PRT; 886 AA.
 AC 09Q87;
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Major outer envelope glycoprotein gp350.
 OS Human herpesvirus 4 (Hepstein-Barr virus).
 OC Viruses; Herpesviridae; no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=10376;
 RN [1]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=98203772; PubMed=9542645;
 RA Lee W.K., Kim S.M., Shim Y.S., Cho S.G., Park S.H., Kim C.W.,
 RA Park J.G.;
 RT "B-lymphoblastoid cell lines from cancer patients";
 RL In Vitro Cell. Dev. Biol. Anim. 34:97-100(1998).
 RN [2]
 RN SEQUENCE FROM N.A.
 RX MEDLINE=99072166; PubMed=9856346;
 RA Chang S.H., Kim S.H., Lee W.K., Kim H.J., Choi S.H., Park J.H.,
 RA Jang H.S., Chung G.H., Kwon T.H., Kim D.H., Yang M.S., Jang Y.S.;
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF172332; AAD51697.1; -;
 DR InterPro: IPR002965; P. rich extensin.
 DR PRINTS; PRO1217; PRICHEXTENSN.
 KW Envelope protein.
 SQ SEQUENCE 886 AA; 92487 MW; F4BCA573B8E29A48 CRC64;
 Query Match 3.9%; Score 132; DB 12; Length 886;
 Best Local Similarity 20.5%; Pred. No. 0.92;
 Matches 139; Conservative 66; Mismatches 268; Indels 206; Gaps 29;
 QY 48 NPTAAADLAQSIDANTVTPPANISSMPEFRNNAKKIDSDSIQWYFYLDPAQATES 107
 DB 169 NITAVR--AQGLD---VTLPLSLPTSADSNSS-----VKTEMG---NEID---IEC 211
 QY 108 ARAVGEYSKIPDGLVPSVDAEIRE-----LYNEPCPVYTDVSVPLDGRQMSLSIFSP 160
 DB 212 IMEDGEISQVLPQDNKFNITCSGYESHVPSGILSTSPVAT--PIPGNYAVASLRITPR 269
 QY 161 PMFR-----TAYVAANVENKMSLD-----VYNDLI-----EMLNLADWRYVDSE 203
 DB 270 PVSRLGNNSILTVFYSGNGPKASGDYCIQSNIVPSDEIPASQDMPTNTTDTIYVGDN- 328
 QY 204 QWINEFTNDTYVYRIRVLRPTVDPTBEGLVTVSVDRLYKATICEANMPTLVDOGFW 263
 DB 329 -----ATYGVPMVTISR-----DANSPVYTVTAFF 352
 QY 264 -----IGQYALPT----- 273
 DB 353 AMPNNTETDFKCKMTLSTGTPSGCENISGAFANRTFDITVSGIGTAPKTLITTRATNA 412
 QY 274 -----SLPQYDVSBVAALHTLTFARPSSAAALAFVWAGLPQGTAPAGTAPMEQAS 324
 DB 413 TTTTHKVFESKABESGTTSTPTLWTPPAVNTT-----TGLPSSGTHVPTNLTA--PTS 463
 QY 325 SCGYLTWRHNGTTFPAGSVSYVLPBGFALERYDPNDG-----SWTDFASAGDTVTRQ 377
 DB 464 TGPTVSTADVTSTPTAGTISGASP--VTPRSPRNGTRESKAPDMTSPISATTTTPRNA 520
 QY 378 VAVDVRVVT-----NPPAGGSAFTTVVPPSNAYNTT-----VFRNTLLSTRPS 423
 DB 521 TSPTPAVTTPTPNATSTPLGKTSPTSAVTTPTPNATSPFAVTTPTPNATITPLTGKTSPT 580
 QY 424 SRRLELPMPPADGQGVVANNPKIEGLKETTLCGYLVHSGRRRPVQLTPASSFGAVSPN 483

```

DB      501 S-AATTPPNATSPVGETSPKANT--NHTIG-----GTSETPVVTSPPKNAISAVTTG 632
QY      404 NPGVERTDLPDYTIR-DSFDQNNSTAVAHFRSLSHSCSIYTKYQMEGVNTVTFPG 542
DB      633 Q--NHITSSSTSMELRPSISSETLSPTSD-NSMSHMLPLLSAHTGENTQV--TPAS 688
QY      543 QPAAAGLKNBEIILCLADDLARLTGVPAT-----DNFAAVASAPAAKMLSSVLKSE 595
DB      609 TSTH-----HVSSTSPAPRPGTTSQASGPNSTSTKPGEVNTTKGTTPPKN 734
QY      596 ATTSIKSVGETAVGAOS 614
DB      735 ATSPAPSGOKTAVPTVTS 753

```

RESULT 9

```

OS6854 PRELIMINARY; PRT; 856 AA.
ID 056854
AC 056854
DT 01-JUN-1998 (TREMBlrel. 06, Created)
DT 01-JUN-1998 (TREMBlrel. 06, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE GP340.
GN BLF1.
OS Human herpesvirus 4 (Epstein-Barr virus).
OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
OC Gammaherpesvirinae; Lymphocryptovirus.
OX NCBI_TaxID=10376;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M81;
RX MEDLINE=97082049; PubMed=8923292;
RA Mackett M., Cox C., Pepper S.D.V., Lees J.F., Beverley B.A.,
RA Wedderburn N., Arrand J.R.;
RT "Immunization of common marmosets with vaccinia virus expressing
RT Epstein-Barr virus (EBV) gp340 and challenge with EBV.";
RL J. Med. Virol. 50:263-271 (1996).
DR EMBL; X99106; CAA67558.1;
DR InterPro; IPR002965; P rich_extenpn.
DR PRINTS; PR01217; PRICHEXTENSN.
SQ SEQUENCE 856 AA; 89634 MW; E34B581AF1071CDD CRC64;

```

Query Match 3.9%; Score 131.5; DB 12; Length 856;
 Best Local Similarity 20.8%; Pred. No. 0.95;
 Matches 120; Conservative 51; Mismatches 190; Indels 217; Gaps 23;

```

QY      48 NETAADLAOSLDANTVTFPANISSMPEFRMA-----KGKIDLSDSIGMYFKYLDP 101
DB      169 NITAVR--AQGLD--VTLPLSLPTSAQDSNFSYKTMQLGNEIDI----- 209
QY      102 AGATESARAVGEYSKIPDGLVKSVDABIRE-----IYNECEPVTVDSVPLDGRQMS 154
DB      210 ---ECIMEDGEISQLPEDNKFNTICSGESHVPSGGLTSTSPAT--PIPGVYANS 263
QY      155 LSTFEPFMR-----TAYVAVANENKEMSLD-----VNDLI---EMLNLADMR 197
DB      264 LELTRPVSRFLGNNISILVYFSGNGPKASGGDYCIQSNIVFSDEIPASQDMPTVTTDT 323
QY      198 YVVDSEQMINFTNDTTYVYRIRVLRPTVDPDTEGLVTVSDYRLTYAICENMPTL 257
DB      324 YVGDN-----ATYSVPMVTSR-----DANSFNV 346
QY      258 VDQGFV-----IGQYALPT----- 273
DB      347 TYTAWAMPNNTETDPCKKMTLTSGPSCENISGAFAFNRPFDITVSLGTAPKTLIT 406
QY      274 ---SLPYDVSEAYALTLTPARSSAALAFVWAGLPQGGTAPAGV- 317
DB      407 RTATNATTTAKVITSKAPESSTTSPTLNTTGFADPNITTTGLPSS--THVPTMLTAPASTG 465
QY      318 -----PAMEQSSGGYLTW-----RHNGT-----TFPAGSVSVLPBEGF----- 351

```

```

DB      466 PTVSTADVTSPTPACTTSGA6PVTPSPSPRNDNGTSKAPDMKTSPTSAVTTPTNGTSP 525
QY      352 ALERYDENDGS---WTDPASAGDTVTFROVADEVVVTNNPAG----- 391
DB      526 AMTTPPNATSPITLGKTSPTSASVTTPTPNATSPPAVTTPPNATSPVGETSPQANATN 585
QY      392 ---GSAAPTPTVRPSPNATNTVFRNTLTETRS6SRRLBPMFPADPGQVANNPKIEQ 448
DB      586 HTLGSTPTPVVTSPPKNAISDV---TTQOHNRSTSSITSSMLRPSIPETISHMPLLTS 642
QY      449 SLKRTLCGYLVH8KMPNVPQLTSPAS-SFGAVSFNNP 485
DB      643 A-----HPTGENTQVTPASISTHVSSTSP 669

```

RESULT 10

```

OS6871 PRELIMINARY; PRT; 1284 AA.
ID 08671
AC 08671
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE Conserved hypothetical protein.
GN MYB6970.
OS Mycoplasma penetrans.
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=28227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF-2;
RX MEDLINE=22354719; PubMed=1246555;
RA Saeki Y., Ishikawa J., Yamashita A., Oshima K., Kenri T., Furuya K.,
RA Yoshino C., Horino A., Shiba T., Saeki T., Hattori M.;
RT "The complete genomic sequence of Mycoplasma penetrans, an
RT intracellular bacterial pathogen in humans.";
RL Nucleic Acids Res. 30:5293-5300 (2002).
DR EMBL; AP004173; BA044489.1;
KM Hypothetical protein; Complete proteome.
SQ SEQUENCE 1284 AA; 141122 MW; CA81CD58A01F7DED CRC64;

```

Query Match 3.9%; Score 131; DB 16; Length 1284;
 Best Local Similarity 18.5%; Pred. No. 2;
 Matches 115; Conservative 72; Mismatches 212; Indels 222; Gaps 28;

```

QY      54 QDLAOSLDAN--TTFPANISSMPEFRMAKGKIDLSDSI-----GWYFKY-- 98
DB      73 QTIADDLILNMLYKLSQTSNVSFNSYNN--ORAQINEDYKDLVDSYKKTGSSMALRQO 131
QY      99 --LDPAGT8ARAVGEYSKIPDGLVKSVDABIREIREIYNECEPVTVDSVPLDGRQMS 156
DB      132 EYLDVVGQTEDA-----YKQ-----KMSWATS 155
QY      157 IFSPMPTAYAV--AN---VENKESLDVNDLIEMNLADMRVYVVDSEQMINFTND 211
DB      156 QPETYLPRTVDLTETDANSNIVDSQSLSKTVSD-----YALQSQKKEAVP 204
QY      212 TTYVYRIRVLRPTVDPDTEGLVTVSDYRLTYKAIT-----CEANMPTLVDOGFVI 264
DB      205 TT-----NTNDYDK8YVQFMQYIMDYQVLENPVYVDSMLK 241
QY      265 GQGYALTPTSLEPYVSEAYALHTLTTPARSSAALAFVWAGLPQGGTAPACTPAMEQAS 324
DB      242 YG--TPTQ-----GIIDYTFASSTTTTENDS 266
QY      325 SGGYLTMRHNGTTPPAGSVSVLPBEGFALERYDPDQSWTDPASAGDTVTFROV----- 378
DB      267 SDDSSSDSDSGSTTANNAGNTIYP-----YFGNDSSTSGSATGTLTKPANFNANS 319
QY      379 AVDEVVY---TNNPAGGSAPTFTYRVPPSNATNTVFRNTLTETRPSRRLBEMPPAD 435
DB      320 ATDTNVTLPSTNGSGSGSGGT--QVKIIPKSYDNTNGY-----GLANTINSYTD 367
QY      436 FQGT--VANNPKIEGSLKETLTCYLVH8KMPNVPQLTSPASSF-----GAVSFNNPQVE 488

```

```

Db 368 DSSVYLAKNTSIYSDL-----YIFFAASSTLFRNNGITGNTNPF 410
QY 489 RTDLDLP---YTGIRDSFDQNSTAVAFRS-----LSHSCSIYTKYQG---WEGVT 535
Db 411 YSADATNSSTHYGI-DSIDSNIQKIGITINNNDPDTTGLDITKQFVSQTDIFSGSS 469
QY 536 NWMPF---GQFAHAGLKNBEILCLADDLATRLTGYTPATDNF-----AAVSAF- 583
Db 470 GKVEYLDLSSVYNQIITKQKLNGLRDNKLYVDSPFIPSDNLTNFMLLNSAGVAIT 529
QY 584 -----AANMLSVLKSZ 595
Db 530 IDGYIFNGSSNGTSTLSVSE 550

RESULT 11
Q9RY11 PRELIMINARY; PRT; 597 AA.
AC Q9RY11;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DE 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE Hypothetical protein DR0143.
GN DR0143.
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC Deinococcaceae; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI.
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Halt D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zaleski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans RI."
RL Science 286:1571-1577(1999).
DR EMBL; AE001876; AAF09732.1; -.
DR TIGR; DR0143; -.
DR InterPro; IPR001434; DUF11.
DR InterPro; IPR001917; NHTransf_2.
DR Pfam; PF01345; DUF11.1.
DR TIGRFAMs; TIGR01451; B ant repeat_4.
DR PROSITE; PS00599; AA_TRANSFERRIN_CLASS_2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 597 AA; 60604 MW; B9A61442EC848279 CRC64;

Query Match 3.9%; Score 130; DB 16; Length 597;
Best Local Similarity 20.6%; Pred. No. 0.63; Indels 172; Gaps 25;
Matches 128; Conservative 67; Mismatches 253;
QY 54 QDLAOSLDANTVTFPANISSMPEFR---NMAKGKIDLDSDISIGMYFYLDPAQAT---- 105
Db 4 RDYDLSLPAAFGGQWVSTSPNQAVCNVSSGGLISNNSYNSLSTRESSGSGATYPPA 63
QY 106 ----ESKAVGEYSKIDGLYKESVDVAIRE--IYNEBCP--VTTDVSVPLDGQMSLSI 157
Db 64 PTAGATWMSVPGNITIPAGTCRFSDVTLPPAGTYKKVIVPANIITSGGNDADGQASITA 123
QY 158 FSPFMPRTAY-----VAVANVENKESLDV-VNDLLEMLNLTADWRYVV 200
Db 124 PACGTVTKAVSPKSLGDGKATSLTTLTNNSSPAVSLTAPLTDNIG--NGLEITGVTT 181
QY 201 DSEQWINTDTTYTVYRIRVLRP-----TYDVPDPTGLVTVSDYRLTYKAITCEAMP 255
Db 182 SCPTANAFSGTITTYPSGATLNPCTITATVRSATAG-----SYF 222

```

```

QY 256 TLVDQG---FWIGGQYA-----LTPTSLPGVDYSEAVA-----LHFLFARPSA 297
Db 223 NTSAGALOTTGNNANNAASDITLVTSTFTLITTKHASQNPFAAGCTGTTVTVSNASGA 282
QY 298 AALAFWAGLPQGGTAPAGTAPWBOASGGGYLTWRNHGTTTPAGSVSYVLPEGFALERYD 357
Db 283 AA-----TSGALSLDDLPSGMSFVSVT 305
QY 358 PN-DGSWTDPASAGDTYTFRQVAVDEVVTNNPAGGSAFTYFRVPPSNAYTNVFRNT 416
Db 306 TTAGSGFRTPASGAT-----GRVDWTFPTSTPLAAGQSILFTVYVAVATVANAGATLTN 360
QY 417 LLETRESRLLELPMPPAD-FGQTVANNPKIEBSLKETIGCTLVHKNENPVQLT--- 472
Db 361 YASVGGGGDDVPLPTPGATCTGSCASDPTTV-----NRTQLTLRK 402
QY 473 --PASRGAVSFNNPQYERTDLPDYTGIRDSFDQNST-AVAFRLSHSCSIYTKYQ 529
Db 403 EFPQGAAGPRGFGN-----YDVATI--TISQSSITLTATTSIANPQVATDT-- 449
QY 530 GWBGVTNN-TTPGQFAHAGLKNBEILCLADDLATRLTGYTPATDNFAAASAPAAAM- 587
Db 450 -----INTTPANTYLRREVLRNDAVFSGPDSYDSR-----YCTNATTSSTVTMPNSS 498
QY 588 -LSVYLKSEATSSIIKSVE 606
Db 499 GMSFTLTPQAGDITITCSVD 518

RESULT 12
Q95S10 PRELIMINARY; PRT; 466 AA.
AC Q95S10;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE H102328P.
GN DSCAM OR CG17800.
OS Drosophila melanogaster (fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Berkley;
RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA Champe M., Chavez C., Dorsett V., Farfan D., Friese R., George R.,
RA Gonzalez M., Guerin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA Nuno J., Paclet J., Paragas V., Park S., Phouanavong S., Wan K.,
RA Yu C., Lewis S.B., Rubin G.M., Celisner S.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY061009; AL28557.1; -.
DR FlyBase; FBgn0033159; Dscam.
DR InterPro; IPR006034; Asp/Glutamase.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR007110; IG_IIIc.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; fn3; 3.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00408; IGc2; 1.
DR PROSITE; PS00144; ASN_GLN_ASR_1; 1.
DR PROSITE; PS50835; IG_LIKB; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 466 AA; 50943 MW; 7056A89B955CC2B CRC64;

Query Match 3.8%; Score 127; DB 5; Length 466;
Best Local Similarity 20.7%; Pred. No. 0.77;
Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;
QY 231 TEGIVRVSDYRLTYKAITCEAMP TLVDQGFWI-----GGQYALTPTSLPGVDYSE 283

```

```

Db      71 TETKQKPHYOMSEATELEKPKP---YEFVWVASTTIGEGQGSKSIVAMPSPQVPAK 126
Qy      284 YA-----LHTLTPARSSAALAFVWAGLPQGGTAPAGTPAMEQASGGYLTRHNGTTP 339
Db      127 IASPDOTPATFKEDAKMPCLA---VGAPO-----PBITWIKINGVER 166
Qy      340 AGSVSYVLPEGFAL-----ERYDPNDGS-WTPASAGDTVTFRQVAV-----380
Db      167 ANDRRKVLPPDGSLLIKSVNRQDAGDYSCHAENSIAKDSITHKLIYALAPQSPHVTLSATT 226
Qy      381 -DEVVVTNNPAGGSGAP--FTVRVPPS-----NAYTN 410
Db      227 TDLTVTKLKPHEGDPAHLGHTLHKPBFGEWETSEVSVDQKHNIIGLCSRRQVYA- 285
Qy      411 TVFRN-----TLLETPSSRRLELPMPPADFGQTVANNPKIEGSLKETLGC-----YL 459
Db      286 TGFNNIGAGEASDIINTRTKGQKPKLPKPR-FIVSSNSVSLHFKAWMDG-GCPMSHFV 343
Qy      460 VHSKKRNP-----VFQUTPASFG--AVSFNNPGY-----487
Db      344 VSKKRDOQIEMNQISNNVKNPNVYVLDLEPATWNLRTAHNSAGFTVAEYDPATLTVT 403
Qy      488 ----ERTDLPD 495
Db      404 GGTIAPSRDLP 415

```

RESULT 13

```

ID      095TGO      PRELIMINARY;      PRT; 1212 AA.
AC      095TGO;
DT      01-DEC-2001 (TREMBlrel. 19, Created)
DT      01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT      01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE      SD09407P.
GN      DSCAM OR CG17800.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_Taxid=7227;
RN      [1]

```

```

RP      SEQUENCE FROM N.A.
RA      Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA      Champe M., Chavez C., Dorsett V., Farfan D., Frisbe R.,
RA      Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA      Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,
RA      Yu C., Lewis S.E., Rubin G.M., Celniker S.,
RA      Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; A058789; AAL14018.1; -
DR      FlyBase; Fgn0033159; Decam.
DR      InterPro; IPR006034; Asp/Glutamase.
DR      InterPro; IPR003961; PN.III.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003598; Ig_G2.
DR      InterPro; IPR003006; Ig_MHC.
DR      Pfam; PF00041; fn3; 6.
DR      Pfam; PF00047; Ig_2.
DR      SMART; SM00060; FN3; 6.
DR      SMART; SM00408; ICG2; 1.
DR      PROSITE; PS00144; ASN_GLM_ASE_1; 1.
DR      PROSITE; PS00835; IG_Like; 2.
KW      Immunoglobulin domain.
SQ      SEQUENCE 1212 AA; 134066 MW; 6D9061E2EPBB5C CRC64;

```

Query Match

```

Best Local Similarity 3.8%; Score 127; DB 5; Length 1212;
Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;

```

```

Qy      231 TEGLVTVSDYRLTYATTCANMPTLVDOGFWI-----GGQVALTPTSLPOVDVSEA 283
Db      455 TETKQKPHYOMSEATELEKPKP---YEFVWVASTTIGEGQGSKSIVAMPSPQVPAK 510

```

```

Qy      284 YA-----LHTLTPARSSAALAFVWAGLPQGGTAPAGTPAMEQASGGYLTRHNGTTP 339
Db      511 IASPDOTPATFKEDAKMPCLA---VGAPO-----PBITWIKINGVER 550
Qy      340 AGSVSYVLPEGFAL-----ERYDPNDGS-WTPASAGDTVTFRQVAV-----380
Db      551 ANDRRKVLPPDGSLLIKSVNRQDAGDYSCHAENSIAKDSITHKLIYALAPQSPHVTLSATT 610
Qy      381 -DEVVVTNNPAGGSGAP--FTVRVPPS-----NAYTN 410
Db      611 TDLTVTKLKPHEGDPAHLGHTLHKPBFGEWETSEVSVDQKHNIIGLCSRRQVYA- 669
Qy      411 TVFRN-----TLLETPSSRRLELPMPPADFGQTVANNPKIEGSLKETLGC-----YL 459
Db      670 TGFNNIGAGEASDIINTRTKGQKPKLPKPR-FIVSSNSVSLHFKAWMDG-GCPMSHFV 727
Qy      460 VHSKKRNP-----VFQUTPASFG--AVSFNNPGY-----487
Db      728 VSKKRDOQIEMNQISNNVKNPNVYVLDLEPATWNLRTAHNSAGFTVAEYDPATLTVT 787
Qy      488 ----ERTDLPD 495
Db      788 GGTIAPSRDLP 799

```

RESULT 14

```

ID      08MKM7      PRELIMINARY;      PRT; 2016 AA.
AC      08MKM7;
DT      01-OCT-2002 (TREMBlrel. 22, Created)
DT      01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DT      01-MAR-2003 (TREMBlrel. 23, Last annotation update)
DE      CG17800-PC.
GN      DSCAM OR CG17800.
OS      Drosophila melanogaster (Fruit fly).
OC      Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC      Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC      Ephydroidea; Drosophilidae; Drosophila.
OX      NCBI_Taxid=7227;
RN      [1]

```

```

RP      SEQUENCE FROM N.A.
RA      STRAIN-Berkeley;
RA      MEDLINE-20196006; PubMed=10731132;
RA      Adams M.D., Celniker S.B., Holt R.A., Evans C.A., Gocayne J.D.,
RA      Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA      George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA      Sutton G.G., Morten J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA      Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
RA      Wan K.H., Doyle C., Baxter B.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA      Abail J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA      Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA      Beeson K.Y., Benoe P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA      Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,
RA      Burris K.C., Busan D.A., Butler H., Cadiou E., Center A., Chandra I.,
RA      Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA      de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA      Dodson K., Dou L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA      Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA      Flosser C., Gabrielian A.B., Garg N.S., Gelbart W.M., Glasser K.,
RA      Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA      Harris N.L., Harvey D., Heiman T.O., Hernandez J.R., Houck J.,
RA      Houlton K.A., Howland T.J., Wei M.-H., Ibeigwan C.,
RA      Jaisli M., Kalush F., Katpen G.H., Ke Z., Kinnison J.A., Ketchum K.A.,
RA      Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA      Lasio P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA      Liu X., Mated J., McIntosh T.C., McLeod M.P., McPherson D.,
RA      Merkulov G., Milshina N.V., Modarres C., Morris J., Moshrefi A.,
RA      Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA      Nelson D.R., Nelson K.A., Nixon K., Nusslein D.J., Pacleb J.M.,
RA      Palazozo M., Pitsman G.S., Pan S., Pollard J., Puti V., Reese M.G.,
RA      Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA      Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,

```

RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RT "The genome sequence of *Drosophila melanogaster*,"
 RL Science 287:2185-2195 (2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Goeys J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barzon J., An H., Baldwin D., Barzon J., Beeson K.Y., Busam D.A.,
 RA Carlson J.W., Center A., Champe W., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorsett V., Doup L.B., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise B., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Paclet J., Paragas V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.,
 RT "Sequencing of *Drosophila melanogaster* genome,"
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Miya S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochick S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
 RA Clump M., Drysdale R., Emmert D., Frise B., de Grey A., Harris N.,
 RA Kronmiller B., Marshall B., Milburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield B.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.,
 RT "Annotation of *Drosophila melanogaster* genome,"
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter J.C.,
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA FlyBase;
 RL Submitted (SEP-2002) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AE003841; AAM6884.1; -
 DR FlyBase; FBgn0031159; Decam.
 DR InterPro; IPR006034; Asp/Glutamase.
 DR InterPro; IPR003961; FN_III.
 DR InterPro; IPR003599; IG.
 DR InterPro; IPR007110; IG-1like.
 DR InterPro; IPR003598; IG_C2.
 DR InterPro; IPR003006; IG_MHC.
 DR Pfam; PF00041; fn3; 6.
 DR Pfam; PF00047; Ig; 9.
 DR SMART; SM00060; FN3; 6.
 DR SMART; SM00409; IG; 10.
 DR SMART; SM00408; IGC2; 9.
 DR PROSITE; PS00144; ASN_GLN_ASE_1; 1.
 DR PROSITE; PS00835; IG_LIKE; 10.
 KM Immunoglobulin domain.
 SQ SEQUENCE 2016 AA; 222168 MW; C3E07ADEBBA4A0F CRC64;
 Query Match 3.8%; Score 127; DB 5; Length 2016;
 Best Local Similarity 20.7%; Pred. No. 8;
 Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;
 QY 231 TEGLVRTVSDYRLTKATTCANMPTLVQGFMT-----GGQYALPTSLIPQYVSEA 283
 DB 1259 TETKQKPHYQMSFEATELEKNKP-----YEFWVASTTIGGQGSIVAMPDQVPAK 1314
 QY 284 VA-----LHTLTPARSSAAALAFVWAGLPQGGTATAGTPAWQASGGLVTRNHTTFFP 339

DB 1315 IASPDFTFTAFKEDAKMPCLA---VGAPQ-----PEITMKIGVERFS 1354
 QY 340 AGSYSTYLPSPGFPAL-----ERYDPDGS-WTDPASAGDTVTRQYAV----- 380
 DB 1355 ANDMRVLPDSSLIKSVNRQDADYSCHEANSIAKOSITKLLVLAPOSPTHLSATT 1414
 QY 381 -DEVVNNNPAGGSGAPT--FTVAVPS-----MYTN 410
 DB 1415 TDALTFLKLPHEGDTPLHGTLLTKYKBPGEWETSEVSDQKINISGLCGSRQYVA- 1473
 QY 411 TVFRN-----TLLETRPSRRRLPMPADPGQTVANNPKIRQSLKETLGC-----YL 459
 DB 1474 TGFNNIGAGSASDILNTRTKQKPLPKPR-FLEVSNSVSLPHFKMKDG-GCPMSHFV 1531
 QY 460 VHSKRNPD-----VQOLTYPASSFG--AVSFNPNFY----- 487
 DB 1532 VESKRRQDIEWNQISNNVKNPNNVYVLDLPATVYNLITAHNSAGFTVAEYDPATLTVT 1591
 QY 488 -----ERTRDLPD 495
 DB 1592 CGGTAPSRDLP 1603
 RESULT 15
 OBMK6 PRELIMINARY; PRT; 2016 AA.
 ID OBMK6
 AC 01-OCT-2002 (TREMREL. 22, Created)
 DT 01-OCT-2002 (TREMREL. 22, Last sequence update)
 DT 01-MAR-2003 (TREMREL. 23, Last annotation update)
 DE CG17800-PA.
 GN DSCAM OR CG17800.
 OS *Drosophila melanogaster* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 CC Empidoidea; Drosophilidae; *Drosophila*.
 CX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Goeys J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.P.,
 RA Searle S.M., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Barker B.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Baas A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brockett P., Brockett P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu B., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos K., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.B., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evansgelist C.C., Ferrara C., Ferreira S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glaeser K.,
 RA Glodek A., Gong P., Gorrell J.H., Gu Z., Guan P., Harris K.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Mei M.-H., Ibegwam C.,
 RA Jallil M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Murtulov G., Mikhlin N.V., Moberly C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny J., Moshrefi A.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
 RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun B.,
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,

RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers R.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of *Drosophila melanogaster*.";
 RL Science 287:2185-2195(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Celinker S.E., Adams M.D., Krommiller B., Wan K.H., Holt R.A.,
 RA Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
 RA Barton J., An H., Baldwin D., Banazon J., Beeson K.Y., Buesam D.A.,
 RA Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
 RA Dodson K., Dorett V., Doup L.B., Doyle C., Dresnek D., Farfan D.,
 RA Ferreira S., Frise E., Galle R.F., Gary N.S., George R.A.,
 RA Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
 RA Ibegwam C., Jaitai M., Kruse D., Li P., Mattei B., Moshrefi A.,
 RA McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
 RA Pacleb J., Patagae V., Park S., Patel S., Pfeiffer B.,
 RA Phouanavong S., Piltman G.S., Puri V., Richards S., Scheeler F.,
 RA Stapleton M., Strong R., Svlyskas R., Tector C., Tyler D.,
 RA Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
 RT "Sequencing of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Misra S., Crosby M.A., Mathews B.B., Bayraktaroglu L., Campbell K.,
 RA Hradecky P., Huang Y., Kaminker J.S., Prochuk S.E., Smith C.D.,
 RA Tupy J.L., Bergman C., Bernan B., Carlson J.W., Celinker S.B.,
 RA Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
 RA Krommiller B., Marshall B., Millburn G., Richter J., Russo S.,
 RA Searle S.M.J., Smith E., Shu S., Smutnick F., Whitfield E.,
 RA Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.B.;
 RT "Annotation of *Drosophila melanogaster* genome.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Adams M.D., Celinker S.E., Gibbs R.A., Rubin G.M., Venter J.C.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Flybase;
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AEO03841; AAM6885.1; -.
 DR Flybase; FBgn003159; Decam.
 DR InterPro: IPR006034; Asp/Glutamase.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR003599; IG.
 DR InterPro: IPR007110; IG-like.
 DR InterPro: IPR003598; IG_c2.
 DR InterPro: IPR003006; IG_MHC.
 DR Pfam; PF00041; fn3; 6.
 DR Pfam; PF00047; Ig; 9.
 DR SMART; SM00060; FN3; 6.
 DR SMART; SM00409; IG; 10.
 DR SMART; SM00408; IGc2; 9.
 DR PROSITE; PS00144; ASN_GLN_ASE_1; 1.
 DR PROSITE; PS50835; IG_LIKE; 10.
 KW Immunoglobulin domain.
 SO SEQUENCE 2016 AA; 222101 MM; DF36AD62BA36FE41 CRC64;
 Query Match 3.8%; Score 127; DB 5; Length 2016;
 Best Local Similarity 20.7%; Pred. No. 8;
 Matches 77; Conservative 45; Mismatches 116; Indels 134; Gaps 18;
 QY 231 TEGLVTVSDRYRTYKATICEANMPTLVGFWI-----GGQVALTPTSLPQYDYSBA 283
 DB 1259 TETKTKQVPHYQMSFEATLEKNKP-----YEFWVTASTTIGEGQGSKIYAPSPDOVPAPK 1314
 QY 284 YA-----LHTLTTPARPSAALAFVWAGLPQGTAPAGTPAWEQASSGGYLTWRHNGTTFP 339
 DB 1315 IASFDDTFTATFKEDAKMPCLA---VGAPO-----PEITWKIKGVFES 1354
 QY 340 AGSVSYVLBEFGAL-----ERYDPNDS-WTDFASAGDTTFRQYAV----- 380

DB 1355 ANDRRKVLPGSLLTIKSVNRQAGYSCHAENSIAKDSITHKLTIVLAPQSPHVTLSATT 1414
 QY 381 -DEVVVTNNPAGGSAFT--FTVRVPS-----NAYTN 410
 DB 1415 TDAIVTKLKHGSDTAPLHGTYLTKPBERGEWETSBSVDSQKNIIGLLCGSRQYVA- 1473
 QY 411 TVFRN-----TLLETSPSRRLBLPMPPADFGQTVANNPKIBOGLKETLGC---YL 459
 DB 1474 TGFNNIGAGASDIINTRTKGQKPLPEKFR-FIVSSNSVSLHFKAWKDG-GCPMSHFV 1531
 QY 460 VHSKRNP-----VFQITPSSFG--AVSFRNPGY----- 487
 DB 1532 VESKRRDQIEMNQISNNVVPDNNYVLDLEPATWNLRTANNSAGFTVARYDPATLTVT 1591
 QY 488 ----ERTDLPD 495
 DB 1592 GGTIAPSRDLP 1603

Search completed: January 15, 2004, 16:52:07
 Job time : 33.869 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 15, 2004, 16:47:33 ; Search time 11.0081 Seconds
(without alignments)
2486.821 Million cell updates/sec

Title: US-09-991-262-50

Perfect score: 3374
Sequence: 1 MGNAGVVASQSPHNRGRTRNV.....GKIAARVRRARRARRARRAN 647

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*\n2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*\n3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*\n4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*\n5: /cgn2_6/ptodata/2/1aa/PTCUS.COMB.pep:*\n6: /cgn2_6/ptodata/2/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3374	100.0	647	3	US-08-485-355B-50
2	3374	100.0	647	3	US-09-194-613-3
3	3033	89.9	675	3	US-08-485-355B-52
4	289	8.6	634	3	US-09-194-613-6
5	139.5	4.1	1231	4	US-09-071-035-420
6	139.5	4.1	1265	4	US-09-071-035-418
7	134.5	4.0	3892	4	US-09-328-352-5503
8	123.5	3.7	878	4	US-09-556-706B-2
9	120	3.6	932	4	US-09-071-035-416
10	120	3.6	969	4	US-09-071-035-414
11	119.5	3.5	907	3	US-08-783-774-2
12	119.5	3.5	907	4	US-09-328-559A-1
13	119.5	3.5	907	5	PCT-US95-04611A-19
14	116	3.4	641	4	US-09-687-538B-8
15	116	3.4	1004	4	US-09-268-347-30
16	113.5	3.4	551	2	US-08-793-229-32
17	113.5	3.4	551	3	US-09-285-957-32
18	113.5	3.4	658	4	US-09-328-559A-2
19	113.5	3.4	816	1	US-07-731-157A-4
20	113.5	3.4	816	1	US-08-229-444B-2
21	113.5	3.4	816	2	US-08-541-780-4
22	111	3.3	20	3	US-08-485-355B-24
23	109	3.2	780	4	US-09-198-452A-920
24	109	3.2	1302	1	US-08-233-537-2
25	108	3.2	591	3	US-09-199-290-7
26	107	3.2	775	1	US-07-603-133B-12
27	106	3.1	618	3	US-09-199-290-34

28	106	3.1	1138	1	US-08-323-474-8	Sequence 8, Appli
29	106	3.1	1138	2	US-08-469-537A-98	Sequence 98, Appli
30	106	3.1	1138	2	US-08-220-240A-5	Sequence 5, Appli
31	105.5	3.1	1611	2	US-08-804-227C-5	Sequence 2, Appli
32	102.5	3.0	497	2	US-08-570-311-2	Sequence 2, Appli
33	102.5	3.0	497	2	US-08-353-485-2	Sequence 2, Appli
34	102.5	3.0	686	3	US-09-306-922-2	Sequence 2, Appli
35	102.5	3.0	2628	2	US-08-570-311-14	Sequence 14, Appli
36	102	3.0	666	2	US-08-737-716-14	Sequence 14, Appli
37	102	3.0	1160	3	US-08-808-599A-24	Sequence 24, Appli
38	101.5	3.0	768	4	US-09-328-352-4402	Sequence 4402, Ap
39	101	3.0	977	4	US-09-252-991A-28261	Sequence 28261, A
40	100.5	3.0	562	2	US-08-687-702-1	Sequence 1, Appli
41	100	3.0	346	2	US-08-687-702-36	Sequence 36, Appli
42	100	3.0	555	3	US-08-588-258B-24	Sequence 24, Appli
43	100	3.0	555	3	US-08-460-505-24	Sequence 24, Appli
44	100	3.0	555	5	PCT-US96-08295-24	Sequence 24, Appli
45	100	3.0	594	4	US-09-468-578-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-485-355B-50
Sequence 50, Application US/08485355B
Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr Hohnbach Test Albitton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treccartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/US8
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 50:
SEQUENCE CHARACTERISTICS:
LENGTH: 647 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 50:
US-08-485-355B-50
Query Match 100.0%; Score 3374; DB 3; Length 647;

Best Local Similarity 100.0%; Pred. No. 1.5e-311; Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MGDAVASQRPNRRGTRNVRVSNATVTVNGRNRQRRTGRQVSPDNFTAAQDLAQL 60
DB 1 MGDAVASQRPNRRGTRNVRVSNATVTVNGRNRQRRTGRQVSPDNFTAAQDLAQL 60
QY 61 DANTVTPANISSMPEFRNNAKGI DLDSDSIGYFKYLDPAGATESARAVEYSKIPDG 120
DB 61 DANTVTPANISSMPEFRNNAKGI DLDSDSIGYFKYLDPAGATESARAVEYSKIPDG 120
QY 121 LVKSVDAIREIINEECPTVDVSVPLDRQMSLISFSPMFRATAYAVANVENKESL 180
DB 121 LVKSVDAIREIINEECPTVDVSVPLDRQMSLISFSPMFRATAYAVANVENKESL 180
QY 181 DVVNDLIEMLNLDKRVYVDSQWINFNTDITYVIRVLRPTYDVPDTEGLVRYVSD 240
DB 181 DVVNDLIEMLNLDKRVYVDSQWINFNTDITYVIRVLRPTYDVPDTEGLVRYVSD 240
QY 241 YRLTYKAITCEANPPTLVDOGFMIIGQYALTPTSIPQYDVBAYALHTLTPARPSAAL 300
DB 241 YRLTYKAITCEANPPTLVDOGFMIIGQYALTPTSIPQYDVBAYALHTLTPARPSAAL 300
QY 301 AFVWAGLPQGGTAAGTPAGTAMEQASSGGYLTWRHNGTTPAGSVSVLPEGFALERYDND 360
DB 301 AFVWAGLPQGGTAAGTPAGTAMEQASSGGYLTWRHNGTTPAGSVSVLPEGFALERYDND 360
QY 361 GSWTDFASAGDTVTFROVAVDEVVVTNNPAGGSAFTVAVPPSNATNTVFRNTLLET 420
DB 361 GSWTDFASAGDTVTFROVAVDEVVVTNNPAGGSAFTVAVPPSNATNTVFRNTLLET 420
QY 421 RPSRRRLLELPMPPADFGQTVANNPKIEQSLKETLGCYLVHKKMNPVFOULTPASSFQAV 480
DB 421 RPSRRRLLELPMPPADFGQTVANNPKIEQSLKETLGCYLVHKKMNPVFOULTPASSFQAV 480
QY 481 SFNPGYERTDLDPTYGIRDSFDQNMSTAVAHFRSLSHSCSIYTKYTOGEGVNTVNTP 540
DB 481 SFNPGYERTDLDPTYGIRDSFDQNMSTAVAHFRSLSHSCSIYTKYTOGEGVNTVNTP 540
QY 541 FGQFAHAGLKNBEILCLADDLATRLTGVPATDNFAAFAAFAANMLSSVYLKSEATSSI 600
DB 541 FGQFAHAGLKNBEILCLADDLATRLTGVPATDNFAAFAAFAANMLSSVYLKSEATSSI 600
QY 601 IKSVEGTAVGAAGSLATLPGILMSVPGKIAARVARRARRARAAN 647
DB 601 IKSVEGTAVGAAGSLATLPGILMSVPGKIAARVARRARRARAAN 647

```

RESULT 2
US-09-194-613-3
Sequence 3, Application US/09194613
Patent No. 6251654
GENERAL INFORMATION:
APPLICANT: GORDON, Karl H.
APPLICANT: HANZLIK, Terry N.
TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDermott, Will & Emery
STREET: 600 13th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3096
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194,613
FILING DATE: 30-NOV-1998
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:
NAME: Bucca Ph.D., Daniel
REGISTRATION NUMBER: 42,368
REFERENCE/DOCKET NUMBER: 50179-061
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-756-8000
TELEFAX: 202-756-8087
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 647 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-194-613-3

Query Match 100.0%; Score 3374; DB 3; Length 647;
Best Local Similarity 100.0%; Pred. No. 1.5e-311; Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 MGDAVASQRPNRRGTRNVRVSNATVTVNGRNRQRRTGRQVSPDNFTAAQDLAQL 60
DB 1 MGDAVASQRPNRRGTRNVRVSNATVTVNGRNRQRRTGRQVSPDNFTAAQDLAQL 60
QY 61 DANTVTPANISSMPEFRNNAKGI DLDSDSIGYFKYLDPAGATESARAVEYSKIPDG 120
DB 61 DANTVTPANISSMPEFRNNAKGI DLDSDSIGYFKYLDPAGATESARAVEYSKIPDG 120
QY 121 LVKSVDAIREIINEECPTVDVSVPLDRQMSLISFSPMFRATAYAVANVENKESL 180
DB 121 LVKSVDAIREIINEECPTVDVSVPLDRQMSLISFSPMFRATAYAVANVENKESL 180
QY 181 DVVNDLIEMLNLDKRVYVDSQWINFNTDITYVIRVLRPTYDVPDTEGLVRYVSD 240
DB 181 DVVNDLIEMLNLDKRVYVDSQWINFNTDITYVIRVLRPTYDVPDTEGLVRYVSD 240
QY 241 YRLTYKAITCEANPPTLVDOGFMIIGQYALTPTSIPQYDVBAYALHTLTPARPSAAL 300
DB 241 YRLTYKAITCEANPPTLVDOGFMIIGQYALTPTSIPQYDVBAYALHTLTPARPSAAL 300
QY 301 AFVWAGLPQGGTAAGTPAGTAMEQASSGGYLTWRHNGTTPAGSVSVLPEGFALERYDND 360
DB 301 AFVWAGLPQGGTAAGTPAGTAMEQASSGGYLTWRHNGTTPAGSVSVLPEGFALERYDND 360
QY 361 GSWTDFASAGDTVTFROVAVDEVVVTNNPAGGSAFTVAVPPSNATNTVFRNTLLET 420
DB 361 GSWTDFASAGDTVTFROVAVDEVVVTNNPAGGSAFTVAVPPSNATNTVFRNTLLET 420
QY 421 RPSRRRLLELPMPPADFGQTVANNPKIEQSLKETLGCYLVHKKMNPVFOULTPASSFQAV 480
DB 421 RPSRRRLLELPMPPADFGQTVANNPKIEQSLKETLGCYLVHKKMNPVFOULTPASSFQAV 480
QY 481 SFNPGYERTDLDPTYGIRDSFDQNMSTAVAHFRSLSHSCSIYTKYTOGEGVNTVNTP 540
DB 481 SFNPGYERTDLDPTYGIRDSFDQNMSTAVAHFRSLSHSCSIYTKYTOGEGVNTVNTP 540
QY 541 FGQFAHAGLKNBEILCLADDLATRLTGVPATDNFAAFAAFAANMLSSVYLKSEATSSI 600
DB 541 FGQFAHAGLKNBEILCLADDLATRLTGVPATDNFAAFAAFAANMLSSVYLKSEATSSI 600
QY 601 IKSVEGTAVGAAGSLATLPGILMSVPGKIAARVARRARRARAAN 647
DB 601 IKSVEGTAVGAAGSLATLPGILMSVPGKIAARVARRARRARAAN 647

```

RESULT 3
US-08-485-355B-52
Sequence 52, Application US/08485355B
Patent No. 6177075
GENERAL INFORMATION:
APPLICANT: Christean, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in Protecting Plants

NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fiehr Hochbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/485,355B
FILING DATE: 07-Jun-1995
CLASSIFICATION: <Unknown>
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.
REGISTRATION NUMBER: 31,801
REFERENCE/DOCKET NUMBER: A-58631-2/RFT/DSS
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 781-1989
TELEFAX: (415) 398-3249
TELEX: 910 277299
INFORMATION FOR SEQ ID NO: 52:
SEQUENCE CHARACTERISTICS:
LENGTH: 675 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 52:
US-08-485-355B-52
Query Match 89.9%; Score 3033; DB 3; Length 675;
Best Local Similarity 100.0%; Pred. No. 4,1e-279; Indels 0; Gaps 0;
Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 68 PANISSMPEFRNNAKGIKIDSDSISIGWYFKYLPDAGATESARAVGEYSKIPDGLVKESVD 127
DB 96 PANISSMPEFRNNAKGIKIDSDSISIGWYFKYLPDAGATESARAVGEYSKIPDGLVKESVD 155
QY 128 AEIREIYNEECPVVTVSVPLDGRQMSLISFSPMFRTAYAVANVENKESLDVNDLI 187
DB 156 AEIREIYNEECPVVTVSVPLDGRQMSLISFSPMFRTAYAVANVENKESLDVNDLI 215
QY 188 EMLNNLADMTYVDSSEMINPTNDTTYVRIKVLKPTVDVDPDREGVLRVTVSDRLTYKA 247
DB 216 EMLNNLADMTYVDSSEMINPTNDTTYVRIKVLKPTVDVDPDREGVLRVTVSDRLTYKA 275
QY 248 ITCEANMPTLVDSGFMTGGQVALPTSLPQYDVSSEAYALHTLTFARSSAAALAFWAGL 307
DB 276 ITCEANMPTLVDSGFMTGGQVALPTSLPQYDVSSEAYALHTLTFARSSAAALAFWAGL 335
QY 308 POGGTAPAGTPAMEQASSGGYLTWRHNGTTPAGSVSYVLEGFALERYDPNDGSMWTFPA 367
DB 336 POGGTAPAGTPAMEQASSGGYLTWRHNGTTPAGSVSYVLEGFALERYDPNDGSMWTFPA 395
QY 368 SAGDTVTFROVADEVVVTNNPAGGASAPFTTVVPPSNATNTVFENLTLETSSRRL 427
DB 396 SAGDTVTFROVADEVVVTNNPAGGASAPFTTVVPPSNATNTVFENLTLETSSRRL 455
QY 428 ELPMPADPGQTVANNPKIEQSLKETLGCYLVHSKRNPNVFOQLTPASSFGAVSFNNPGY 487
DB 456 ELPMPADPGQTVANNPKIEQSLKETLGCYLVHSKRNPNVFOQLTPASSFGAVSFNNPGY 515

QY 488 ERTDLPDYTGIRDSFDQNMSTVAHFRSLSHSCSIYTKYQMGEGVTVNTPFGQFAHA 547
DB 516 ERTDLPDYTGIRDSFDQNMSTVAHFRSLSHSCSIYTKYQMGEGVTVNTPFGQFAHA 575
QY 548 GILKNEEILCLADDLATRLTGYYPATDNPAVAASAPAAANMLSSVLKSAIISIKVGFT 607
DB 576 GILKNEEILCLADDLATRLTGYYPATDNPAVAASAPAAANMLSSVLKSAIISIKVGFT 635
QY 608 AVGAQSGIAKLPGLMSVPGKIAARVBARBRARRAARAN 647
DB 636 AVGAQSGIAKLPGLMSVPGKIAARVBARBRARRAARAN 675
RESULT 4
US-09-194-613-6
Sequence 6, Application US/09194613
Patent No. 6251654
GENERAL INFORMATION:
APPLICANT: GORDON, Karl H.
TITLE OF INVENTION: MODIFIED SMALL RNA VIRUSES
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: McDermott, Will & Emery
STREET: 600 13th Street, N.W.
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3096
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/194,613
FILING DATE: 30-NOV-1998
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Bucca Ph.D., Daniel
REGISTRATION NUMBER: 42,368
REFERENCE/DOCKET NUMBER: 50179-061
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-756-8000
TELEFAX: 202-756-8087
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 634 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-194-613-6
Query Match 8.6%; Score 289; DB 3; Length 634;
Best Local Similarity 25.7%; Pred. No. 2e-18;
Matches 177; Conservative 72; Mismatches 266; Indels 174; Gaps 39;
QY 35 QRRRTGRVSPPNFRAAA--DLAOSLDANTVTFPANISGM-----PE-----76
DB 31 QRADKTRBKPPRRATAAQTITTTTQKAGSKTSPRSKRIDYQAPAKPNRPRRHPOQR 90
QY 77 --PNNMAKGIKID--SDSIGWYFKYLPDAGATESARAVGEYSKIPDGLVKESVDAR 131
DB 91 SDTRGAKASDDSHSDIKAMTHDYLDDGKYSKISLDG---KIPDGAIPGTCGQFR 147
QY 132 EYNEBECVVTVSVPLDGRQMSLISFSPMFRTAYAVANVENKESLDVYV--DL-----186
DB 148 GTVGARYPGNSLTLLPLDGGTWPPLVHMLPFPFRPLPITTSBTR--VEVTNADLDAFA 205
QY 187 EMLNNLADMTYVDSSEMINPTNDTTYVRIKVLKPTVDVDPDREGVLRVTVSDRLTYKA 242
DB 206 NDV--NRTDWTAT--YSSMAQVGN--VFVWV-----FTBALTDVPPPTGLGVSGLLSYR 257

```

QY 243 LTYKAITCEANMPTLVDOGFMIIGQYALTPISLPOVDVSEAYALMTLT-----PAPPS 295
DB 258 LSSGVTAVFNAPFTLVNOGVAVIAQF--QPDKEHKKENDIVAGTQTGGTLOLQSGSGPN 315
QY 296 SAALAFWAGIPOGTA-PACTPAMEQSSGGYLTWRHNGTTPAG--SVSYLPEGF 351
DB 316 YTLTNT-IGDOVEFGAALPLFTVSMGMPESGQLVFQANLTFDVGNITITITTLPPROS 374
QY 352 ALERYDPDGSWTDPSAG-DTATPROVAVDEVVTVNNPAGGSAPTPTVRVPSPNAYTN 410
DB 375 V-----TGMMQFASNGDTVT-----VD-----AGA-----TVRV----- 400
QY 411 TVFNRTL-LETRPSSRLLE---PMRPADFGQVANNPKIEQSLKETLCCYLVSMM 464
DB 401 ---RSEFGLTGESAGHQLNODSTNDMNPNDAG---NAKTIQFOLTKR--CHYMEASI 451
QY 465 RNPVQLPASPFGAVSFNN-----PQYER--TRD-LPDVGTGRDSDQNM 507
DB 452 R---VANDNATISYGPDEDETDNCGRLHRAIGALQGYHRQQLDRLPSTMG-----M 500
QY 508 STAVAFPSLSHSCSIVTKTYQMEGV-----TVNTPFGQFAHAGLKNBEILCLADD 561
DB 501 STSTPY-----LQGVPTLRSDTGGGEPMPFASATPPKDDVALTVART 544
QY 562 LATRLTGVPATDNFPAAVSAPAAANMLSSVLKSEATSSIIKSVGETAVGAQSGIAKLPG 621
DB 545 WTLDPFAFPERYNGFALFAMVAKTIAQIPR-----YVRS---AAGVANAVTDCIES 594
QY 622 LMSVPGKIAARVARBARR-----RAAR 645
DB 595 ATEVSANSTERRORRARRVGIGIARGAR 623

```

RESULT 5
US-09-071-035-420

```

; Sequence 420, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gail H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071, 035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: A. Anders Brookes
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB369P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 420:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1231 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

```

TOPOLOGY: linear
MOLECULAR TYPE: protein
US-09-071-035-420

Query Match 4.1%, Score 139.5, DB 4, Length 1231;
Best Local Similarity 19.0%, Pred. No. 0.001;
Matches 141; Conservative 94; Mismatches 252; Indels 257; Gaps 31;

```

QY 19 NRVSAATVTVNGRRNRORRGROVSPPDNTAAQDLASLDANTV---TPPNISS- 73
DB 542 NRKVTENFVDNNGAK-----ITPPTGFGGKKTIVTS-DATTFKQAGTLPTDTYTG 591
QY 74 --MPEFRWAKGKIDLSDSIGWYFKYLDPAAGATSSAAVBSKIPGLVFSVDARIR 131
DB 592 GRTYFKGMWYKSKILNTLT-----TYAPSYQVYDNDIDLN 629
QY 132 ELYNEECVAVTVDSVPLD-----GROWSLIFSPMPRTAVAAVANYEN 175
DB 630 VVYREFTVTVPSVDNMFVBEKGAFTPALTFSGKYAAGTSAY--LRTDLVDTSKRN 687
QY 176 -----KENSLLDVNDLIBWLN----- 191
DB 688 GNGQTVSINNQMPLSQBLKTKYNNQOPISATNRLQFNVDKLAIQQLKYVDSIQDLPTA 747
QY 192 ---NLADRRYV-----DSEQMINPTNDTTYV---RIRVLRP 223
DB 748 GSSNLKSYRYVTNNSLVFPDENVAPAEVDLSSSLNTLNPDSDTYSNANNRLPYTHL 807
QY 224 TYD-VPDPTBGLVRTVSD-----YRLTYKAITCEANMPTLVDOGFMIIGQYALTP 272
DB 808 GYSGTFGVNYYLLVWMLFNAKPADSKLYKTKRKQVT-----ENVVD---VNGAKITRP 858
QY 273 TSLPOYDVSEAYALHTLTPAPBSAAL-----AFWAGLPQGTAPGTPAMEQA 323
DB 859 TGFTQ---GNOVPNMSNTF-KYTAAKALPATYTTGKAYTFPGWYKGTKPSTL----- 908
QY 324 SSGGLTWRHNGTTPAGSVSYVLPBGFALERYDNDPSQWMDPASAGDTVTRQVAVDEV 383
DB 909 -----NKTTPTRPAT-----FDQNDMDTMYKKEIPFASVTLTRPKEV 947
QY 384 VVTNPPAGGSAPTFTVVPSPNAYTNVFRNTLLETBPS-SRLRLP--MPPADGQTV 440
DB 948 IDTNINVTITITITSKAP-----LQNLTLKQPNMGAGLITPTEFVTEBGETT 998
QY 441 ANPKIEQSLKETLCTLVHSMKRNPFVQLTTPASSFGAVSF-----NNPGETRDLDP 495
DB 999 KSIP-VNSTLMTBEGV-----PLPNAPVPIGKTVSVAPTRATGKFNVLKAEVVV 1046
QY 496 YTGIRDSPQNMSTAVAHFRSLSHSCSIVTKTYQMEGVTVNTPFGQFAHAGLKNBEI 555
DB 1047 FGIHDSITVDNF-----VRIKPNQOBVVTPTEBGFISVPTFD--FGQGVAGTKQSHL 1098
QY 556 LCLAD-----DLATRLGVYPATDNF-----AAVSAFPAAN 586
DB 1099 KQADVYNGGNTNPFYRIKKTQPNWSLTQSLQSPASATDSLPTARLLLGAAVPSFTNY 1158
QY 587 MLSVVKSE--ATSGIIKSVGETA 608
DB 1159 NQPTLKNVGTTSALSILTANNTA 1182

```

RESULT 6
US-09-071-035-418

```

; Sequence 418, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gail H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville

```

STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/071.035
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 418:
SEQUENCE CHARACTERISTICS:
LENGTH: 1265 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-071-035-418

Query Match 4.1%; Score 139.5; DB 4; Length 1265;

Best Local Similarity 19.0%; Pred. No. 0.0011;

Matches 141; Conservative 94; Mismatches 252; Indels 257; Gaps 31;

QY 19 NVRVATVYVNGRRNRRTGRVSPDNETAAQDLAOSLDANTV---TFPANIS- 73
DB 574 NRKYTEFVDTNGK-----ITPPTGTOGKIVITS-DATYKQAGTLPDYTTG 623
QY 74 -MEFNNMAKGKIDLSDSIGWFKYLDPAGATESARAVEGYSKIDGLVKSVDAR 131
DB 624 GKTKFKGMYKSKSLNLTLT-----TKASVQVTDNDNDLN 661
QY 132 EYNEECPPVTVDSVPLD-----GROMSLISFPMFRAYVAVAVEN 175
DB 662 VVVEEETVTVYVSDNPFVNEKGAFTPALTFSGKYAOSTSAY-LRTDLYDVTSKN 719
QY 176 -----KEMSLDVNDLIEMLN----- 191
DB 720 GNGQYVTSINNGSMPLSQELIKYNNQOPISATNRLOPNVDKLAIDQGLKVSIDDTA 779
QY 192 ---NLADMRVYV-----DSEQMINFTNDTTYV---RIRVLAR 223
DB 780 OSSSLKRYRYVYTNNSLVEFDNVAARVLDLSESLNULNFDSDGYFSNANRLPYTHL 839
QY 224 TYD-VPTPTGGLVTVSD-----YRLTYAIIICEANMPLVUOGFWIGQVATLP 272
DB 840 GYSGTPEVNYLVLFENAKPADRSKLVYKTRQVY-----ENFVY-----VNGAKITAP 890
QY 273 TSLPOYVSEAYVALHTLTFARPSAAL-----AFVWAGLPQGGTAPAGPAMEQA 323
DB 891 TGFQO---GNOVPMNSVTF-KYIAKALPATYTTGSKYVTFQGYKKGKTRSTL----- 940
QY 324 SSGGYLTMRHNGITFPAGSVSYVLBGFALERVDPNDSWTFASAGDTVTFROVADEV 383
DB 941 -----NKTTPTFENAT-----FDGNDMTAMYKEBIPASVLTLPREKV 979
QY 384 VVTNPAAGGSAPFTYVAVPPSNAYTNTVFRNTILTRPS-SRLLELP--MPPADFGQTV 440
DB 980 IDNTNVIWTTTINTSKAP-----LQNLTLKGGPMNSAGLITPTFEVTPPEBETT 1030
QY 441 ANNPKISQSLKETLGCYLVSHKRNPNVFLQTPASSFGAVSF-----NNPGYERTDLPD 495

DB 1031 KSIP-VNLSLMTGCV-----PLPNAVPIGKKVSAFTTRANGKENTVLKAEVV 1078
QY 496 YNGIRDSFDMNSTAAHPRSLSHSCSIYTKYQHGCVTNNVTPPGQPAAGLKKEI 555
DB 1079 FGQIKOSTVDNF-----VAIRPNDSVYPTTSGPISVTFD--FGQVGAOTKQOHS 1130
QY 556 LCLAD-----DLATLTGVYPATDNF-----AAVSAFAN 586
DB 1131 KQADDTYNGNTRNPIYRIKKTQPMNSLTAQSPKASATBSLPNTRLLGLAAVSSPTNY 1190
QY 587 MESSVLSR--ATSSIKSVGETA 608
DB 1191 NQPTLKNVTGTTSAISLFLANNTA 1214

RESULT 7

US-09-328-352-5503
Sequence 5503, Application US/09328352
Patent No. 6562958
GENERAL INFORMATION:

APPLICANT: Gary L. Breton et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5503
LENGTH: 3892
TYPE: PRT
ORGANISM: Acinetobacter baumannii

US-09-328-352-5503

Query Match 4.0%; Score 134.5; DB 4; Length 3892;

Best Local Similarity 21.1%; Pred. No. 0.022;

Matches 146; Conservative 84; Mismatches 306; Indels 157; Gaps 32;

QY 23 SANTVYVNGRRNRRTGRVSP---PDNFTAAQDLAOSLDANTVTFP---ANISMP 75
DB 291 AGNSSVNDQYTLDTLDTAANAVIDPVGNDTDTGTABGSTVTVYYPGSGTKTVVAGP 350
QY 76 BFRNMAKGKIDL-DSDSIGWFKYLDPAAG-----ATESARAVEGYSKIPGLVKSVDAR 129
DB 351 D-GTWYVNPNGLDGDEV--TAVATDPAGTSGPATVYDAVAPALADVLVINDSTPAL 407
QY 130 IREIYNEECVTVYD-----SVPLDGRQMSLSISFPMFR-----TAYVAVANV-- 173
DB 408 TGVNDPTATVYVNVVDGVDYPAVNNGDGTTLADNTLPTLADGPHITVYATDAAGVGT 467
QY 174 ENKEMSLDVNDLIEMLNLMADMRVYVDSQMINFTNDTTYVYRIRVLARTTYVDPDTBG 233
DB 468 DTGVVTVDTA-----APNTAGVTFTIDSVTADNVINASSAAGNVITTVGLKXNPADATN 521
QY 234 LVETVSDYRLTYRAI-----TCBANMP---TLVDQFWIGGOVALTPTSILPOYDSEAY 284
DB 522 TATTVVINGVTTAATYDKTAGTWTVSFSGGLVADADKTIDAKYTFDDAAGNSTVNDT- 580
QY 285 ALHTLTFARPSAALAFVWAGLPQGGTAPAGPAMEQASGGYLTMRHNGITFPAGSVS 344
DB 581 QITLDTAADAAPVI-----DPVNGTDP1-----GTABRGSVTV 615
QY 345 YVLBEGALRYRPNDSQWT---DFASAGTTVTFROVADEVYVYTNPNPAGGSAP-TFTV 400
DB 616 VTYFNGDTATVAGPDGSMGVPFPGLDGDEV-----EALAT-DPAGNPSLPGTATV 666
QY 401 -RYVPS---NATNTVFRNTILTRPSRRLB---LMPPADFGQTVANNPKIBQSL 451
DB 667 DANGPNTDGNFTYVDSLTADNVINASSAGNVYTVGLKXNPADAANTV----- 715
QY 452 KETLGCYLVSHKRNPNVFLQTPASSFGAVSFNNPGYERTD-----LDPYTS--- 498
DB 716 -----VTVVINGQVYATVDSGTAGTWTVSFSGSLTADADKTIDAKYTFDDAAGNS 767

QY 499 -IRSDPQNMSTAVAH-----FRSLSHSCSYTKYQGWEGVNTNTPGCPA 545
 DB 768 SYNDTQTTTITDAPAPINPVNGTDTTGTABPGSTVTVYPDGSTTTVAAGPGXT 827
 QY 546 --HAGLKNEEILCLADLTLRLTGVYPATDNFAAANMSSVLKSEATSSIIKS 603
 DB 828 VNPGLNDQDKYTAITATDPAGNPS--LPGT-----ATVDVAGPR-----TGVNPTVDS 874
 QY 604 V-GETRAGAAQ-SGLAKLPGLIMSVPGKTAARY 634
 DB 875 VTADNVINASEASGNVTGTGVLKNVPADANTV 907

RESULT 8

US-09-556-706B-2
 / Sequence 2, Application US/09556706B
 / Patent No. 6458364
 / GENERAL INFORMATION:
 / APPLICANT: Spaeete, Richard
 / APPLICANT: Jackman, Winthrop
 / TITLE OF INVENTION: NON SPLICING VARIANTS OF GP350/220
 / FILE REFERENCE: 7682-050-999
 / CURRENT APPLICATION NUMBER: US/09/556, 706B
 / CURRENT FILING DATE: 2000-04-24
 / PRIOR APPLICATION NUMBER: 08/783,774
 / PRIOR FILING DATE: 1997-01-15
 / PRIOR APPLICATION NUMBER: 08/229,291
 / PRIOR FILING DATE: 1994-04-18
 / NUMBER OF SEQ ID NOS: 19
 / SOFTWARE: PatentIn version 3.0
 / SEQ ID NO 2
 / LENGTH: 878
 / TYPE: PRT
 / ORGANISM: Virus
 / FEATURE:
 / OTHER INFORMATION: gp350
 / US-09-556-706B-2

Query Match 3.7%; Score 123.5; DB 4; Length 878;
 Best Local Similarity 19.9%; Pred. No. 0.019;
 Matches 138; Conservative 67; Mismatches 256; Indels 231; Gaps 28;

QY 48 NPTAAQDLAQSILATNTTPPANISSMPFRNNAKIKLDDSSISIMRYKIDLPAGATES 107
 DB 169 NITAVR--AQGLD--VTLPLSLPTSADDSNFS--VKTEMLD--NEID--ISC 211
 QY 108 ARAVGEYSKIPDLVFSVDABIRE-----IYNEECPVTVDSVPLDGRQWLSISF 160
 DB 212 IMEDGISQVLPEDNKNFNTCSGYESHVSGGILTSTSPVAT--PIPGGYAVSLRLTR 269
 QY 161 PMFR-----TAYVAVANENKMSLD-----VNDLI-----EWLANLADWRYVDSB 203
 DB 270 PVSRLFNNISILVFPYSGNGPKASGDYCIOSNIVFSDIIPASQDMPTVTTDITVYGDN- 328
 QY 204 QMINFTNDTTYRIRIVAPTYVDPPTBGLVATVSDYKLTAYTICEANMPLVVOGFW 263
 DB 329 -----ATYSVPMTVTS-----DANSPVVTATW 352
 QY 264 -----IGGOYALTP----- 273
 DB 353 AMNNTEPTDKCKWTLTSGPSCENISGAFASNRFPDITVSGLTGAPKTLITTRATNA 412
 QY 274 -----SLPOYDVSEAVYALHTLTPARPSSAALAFVWAGLPQGTAPAGTPAMEQAS 324
 DB 413 TTTTHKVFISKABESTTSTLTNTGTGADENNT-----TGAPSTHVPTNLTA--PAS 463
 QY 325 SGGYLTWRNGTFFPAGSVSYVLPBGFALERVDPNDSW-----TFPAGSDVTFR- 375
 DB 464 TGETVSTADVTSPTPGTTSNASP-----VTSPSPMDNGTESKAPMPTSTSTSPVTP 516
 QY 376 RQAVAVEVVVNNPAGGSAFTFVVPSPNATYNTVFRNTLLTETPSSRLTLPMPAD 435
 DB 517 TPNAJSTPTAVTTPPTNATSPTPAVTTPPTNATS-----PTLQKTSPTS-AVTTTPPN- 569

QY 436 PGQYANNPKIQSLIKETIGCYLVHSGKRPVFPOLTPASSPGAVSFNNPGRTRDLPD 495
 DB 570 -----TSPTLAKTSPPTSAVTP-----TPNATSPPTLAKTSPPTSAVTPPTPN 610
 QY 496 YTG-----IRSDPQNMSTAVA---HFR-----SLSHSCS 522
 DB 611 ATGPTVGETSPQANATNTHTLGSTSPPTVYTSQPKATATAVTTGGHNRPSNETLSPST 670
 QY 523 IYTKYQGWEGVNTNTPPGQFAHAGLKNEEILCLADLTLRLTGVYPATDNFAAANMSS 582
 DB 671 DNSTSHMGENTQY-TPASISTH-----HVSSTSPPRAGTTSQASGPNSSSTSTXP 722
 QY 583 PAANMLSSVLKSEATSSIIKSIGTRAVGAAGS 614
 DB 723 GEVNTKGTTPQNAATSPQAPSGQKTAVPVTVTS 754

RESULT 9

US-09-071-035-416
 / Sequence 416, Application US/09071035
 / Patent No. 6448043
 / GENERAL INFORMATION:
 / APPLICANT: Gail H. Choi
 / TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
 / NUMBER OF SEQUENCES: 496
 / CORRESPONDENCE ADDRESS:
 / ADDRESSER: Human Genome Sciences, Inc.
 / STREET: 9410 Key West Avenue
 / CITY: Rockville
 / STATE: Maryland
 / COUNTRY: USA
 / ZIP: 20850
 / COMPUTER READABLE FORM:
 / MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
 / COMPUTER: HP Vectra 486/33
 / OPERATING SYSTEM: MSDOS version 6.2
 / SOFTWARE: ASCII Text
 / CURRENT APPLICATION DATA:
 / APPLICATION NUMBER: US/09/071,035
 / FILING DATE:
 / CLASSIFICATION:
 / PRIOR APPLICATION DATA:
 / APPLICATION NUMBER:
 / FILING DATE:
 / ATTORNEY/AGENT INFORMATION:
 / NAME: A. Anders Brookes
 / REGISTRATION NUMBER: 36,373
 / REFERENCE/DOCKET NUMBER: PB369P2
 / TELECOMMUNICATION INFORMATION:
 / TELEPHONE: (301) 309-8504
 / TELEFAX: (301) 309-8512
 / INFORMATION FOR SEQ ID NO: 416:
 / SEQUENCE CHARACTERISTICS:
 / LENGTH: 932 amino acids
 / TYPE: amino acid
 / STRANDEDNESS: single
 / TOPOLOGY: linear
 / MOLECULE TYPE: protein
 / US-09-071-035-416

Query Match 3.6%; Score 120; DB 4; Length 932;
 Best Local Similarity 19.5%; Pred. No. 0.046;
 Matches 149; Conservative 95; Mismatches 249; Indels 272; Gaps 45;

QY 2 GAGVAAQRPNHRGRNVRVSGANTVTVNGRNRORT-----GROVSPDNFPAALAO 54
 DB 222 GAVG-GPTSSHQRKRGAG-----PYTYITLKRVTKEFNPAGAGALPAEBGVT---Q 269
 QY 55 DLQASIDANTVTPPANISSMP-----FNNWAKGKI---DLDSDSIGWY----- 95
 DB 270 DKKTIVTGEDPFTTQEG-GTLPERYTGSDOKTYLFRGWKGNKAKPSTLLETTKTPSAVTVYD 328

```

Qy 96 -----FKYLDPAQ-----AT--ESARAVEYSKI 117
Db 329 DNDLHVYERAVKTYTLPARBALFGYVDEQGLINPAKRLSATMGESGATGEMTTF 388
Qy 118 P--DG-----LVKFSVDABIREIYN--BECPVVT-----DVSVPDLGRQMSLISFSPM 162
Db 389 PTIDGIMPAQLKKAIP--QKYYTRPDGTYITVGPQSVSEIPIKYQTISISP-- 442
Qy 163 FRTAYVA-----VAN-----VENKMSLDVNDLIEM-----LNLADMRVYVDSQWIN 207
Db 443 -TTAYTGDKTKYPPNENVRGRIENPD--NIVSSLVGKXAVNLQKATRTAARSYXW 498
Qy 208 FTNDTTYVRI--RVLRPTVDPPTEGLVRTVSQVRLTYKALTCBAMPTLVDOGFWIG 265
Db 499 GPTILYSMSIYSTAGANTNLSTP--DGTIYYILENRVYTHFVDESAGAKITPTGFTQG 557
Qy 266 GQVALTPTSLPQYDVSAAVALHTLTPARPS--SAAALAFVWAGLPQGGTAPAG----- 316
Db 558 NQLVVD-----SENY--VYTVAKALPKIYQAGEKTYIFQGWFKGKTKRATLKTITTP 607
Qy 317 -----TAMEQASSGCYLTVRHNQITTPAGSVSYVLPEGFALERYDPDQSWT 364
Db 608 SFTPTFNDDEDMTAVYQBALPTAELTL-----TGAVDII-----ENGATM 647
Qy 365 DFASAGDTVTRQYAVDEVVVVNNPAGGSAPTFTVRVPSNAYT-----NTVF----- 413
Db 648 DYM-----EALLKNT-----GAAPLTIIKIKPTAYTAAGIAPANTIFVQCTG 689
Qy 414 RNTLETRPSSRR-----LELPMPPAD-----FGQTVANNP----- 444
Db 690 QNT--KAPFVTKQMTTGAGVSIITLDQPLPAGGQKNNLGTAVTGNQGLTADVEVYG 747
Qy 445 KIBESLKETLGCIVLSKRNPN--VFQULTPASFG--AVSFNNPGERTRDLPDY--T 497
Db 748 NFGSLTAKDVRIRKDLQBITSPDGGFISTPTDFGKLAISSGKOQY--GLKKAADYYGN 806
Qy 498 GIRDSP-----DQNMSTAVAHFRSLSHSCSIYTK-----TYQGEVGYTNVNTP 540
Db 807 GTRNPIRLNTSQANSLTQOLSQPKSATDSLPTTRILLGTAAASSTVDYNOPTETRTTP 866
Qy 541 FQGFPAHAGLKNBEILCLADLAT-----RLTG--VYPATDNFA 577
Db 867 LG-----KTSVTYTLTADNTATAVAVANQOFTGSDVYQDLPTFA 903

RESULT 10
US-09-071-035-414
; Sequence 414, Application US/09071035
; Patent No. 6448043
; GENERAL INFORMATION:
; APPLICANT: Gil H. Choi
; TITLE OF INVENTION: Enterococcus faecalis Polynucleotides and Polypeptides
; NUMBER OF SEQUENCES: 496
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/071.035
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; ATTORNEY/AGENT INFORMATION:

```

```

NAME: A. Anders Brookes
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB369P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 414:
SEQUENCE CHARACTERISTICS:
LENGTH: 969 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULAR TYPE: protein
US-09-071-035-414

Query Match 3.6%, Score 120, DB 4, Length 969,
Best Local Similarity 19.5%, Pred. No. 0.049;
Matches 149, Conservative 95, Mismatches 249, Indels 272, Gaps 45;

Qy 2 GDAGVASQRPNNRGTNNRVASANTVTNGRRNRRT-----GQVSPDNFTAAQ 54
Db 259 GDVG-GPTSSHORGTAG-----PYYVYLTGRVTRKFPENPAGAIAPBEGYT---Q 306
Qy 55 DLAGSLANTVTPANISSMP-----FRNAKGL--DIDSISGY----- 95
Db 307 DKTTIVGEDFTTQO--GTLPERYTGSDGKTYLFQGWYKGNAPKSTLETTKTPSYAVTYD 365
Qy 96 -----FKYLDPAQ-----AT--ESARAVEYSKI 117
Db 366 DNDLHVYERAVKTYTLPARBALFGYVDEQGLINPAKRLSATMGESGATGEMTTF 425
Qy 118 P--DG-----LVKFSVDABIREIYN--BECPVVT-----DVSVPDLGRQMSLISFSPM 162
Db 426 PTIDGIMPAQLKKAIP--QKYYTRPDGTYITVGPQSVSEIPIKYQTISISP-- 479
Qy 163 FRTAYVA-----VAN-----VENKMSLDVNDLIEM-----LNLADMRVYVDSQWIN 207
Db 480 -TTAYTGDKTKYPPNENVRGRIENPD--NIVSSLVGKXAVNLQKATRTAARSYXW 535
Qy 208 FTNDTTYVRI--RVLRPTVDPPTEGLVRTVSQVRLTYKALTCBAMPTLVDOGFWIG 265
Db 536 GPTILYSMSIYSTAGANTNLSTP--DGTIYYILENRVYTHFVDESAGAKITPTGFTQG 594
Qy 266 GQVALTPTSLPQYDVSAAVALHTLTPARPS--SAAALAFVWAGLPQGGTAPAG----- 316
Db 595 NQLVVD-----SENY--VYTVAKALPKIYQAGEKTYIFQGWFKGKTKRATLKTITTP 644
Qy 317 -----TAMEQASSGCYLTVRHNQITTPAGSVSYVLPEGFALERYDPDQSWT 364
Db 645 SFTPTFNDDEDMTAVYQBALPTAELTL-----TGAVDII-----ENGATM 684
Qy 365 DFASAGDTVTRQYAVDEVVVVNNPAGGSAPTFTVRVPSNAYT-----NTVF----- 413
Db 685 DYM-----EALLKNT-----GAAPLTIIKIKPTAYTAAGIAPANTIFVQCTG 726
Qy 414 RNTLETRPSSRR-----LELPMPPAD-----FGQTVANNP----- 444
Db 727 QNT--KAPFVTKQMTTGAGVSIITLDQPLPAGGQKNNLGTAVTGNQGLTADVEVYG 784
Qy 445 KIBESLKETLGCIVLSKRNPN--VFQULTPASFG--AVSFNNPGERTRDLPDY--T 497
Db 785 NFGSLTAKDVRIRKDLQBITSPDGGFISTPTDFGKLAISSGKOQY--GLKKAADYYGN 843
Qy 498 GIRDSP-----DQNMSTAVAHFRSLSHSCSIYTK-----TYQGEVGYTNVNTP 540
Db 844 GTRNPIRLNTSQANSLTQOLSQPKSATDSLPTTRILLGTAAASSTVDYNOPTETRTTP 903
Qy 541 FQGFPAHAGLKNBEILCLADLAT-----RLTG--VYPATDNFA 577
Db 904 LG-----KTSVTYTLTADNTATAVAVANQOFTGSDVYQDLPTFA 940

RESULT 11

```

US-08-783-774-2
 ; Sequence 2, Application US/08783774
 ; Patent No. 6054130
 ; GENERAL INFORMATION:
 ; APPLICANT: Spaete, Richard
 ; APPLICANT: Jackman, Winthrop
 ; TITLE OF INVENTION: NON-SPLICING VARIANTS OF
 ; TITLE OF INVENTION: GP350/220
 ; NUMBER OF SEQUENCES: 19
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Pennie & Edmonds
 ; STREET: 1155 Avenue of the Americas
 ; CITY: New York
 ; STATE: NY
 ; COUNTRY: USA
 ; ZIP: 10036/2711
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette
 ; COMPUTER: IBM Compatible
 ; OPERATING SYSTEM: DOS
 ; SOFTWARE: FastSeq Version 2.0
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/783,774
 ; FILING DATE: 15-JAN-1997
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Coruzzi, Laura A.
 ; REGISTRATION NUMBER: 30,742
 ; REFERENCE/DOCKET NUMBER: 7682-037
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-790-9090
 ; TELEFAX: 212-869-8864
 ; TELEX: 66141 PENNIE
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 907 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS:
 ; TOPOLOGY: unknown
 ; MOLECULE TYPE: protein
 ; US-08-783-774-2

Query Match 3.5%; Score 119.5; DB 3; Length 907;
 Best Local Similarity 19.5%; Pred. No. 0.049;
 Matches 136; Conservative 67; Mismatches 274; Indels 219; Gaps 27;
 QY 48 NFTAADLAQSLDANTVTFFPNNISMPFRMAKIDLDSDISIKYFKYIDPAGATES 107
 DB 169 NITAVVR--AAGLD--VTLPUSLPTSADSNFS--VKTEMLD--NEID---IEC 211
 QY 108 ARAVGEYSKIPDGLVFSVDAEIRE-----IYNECPVTVDVSVPLDGRQWSLSIFSF 160
 DB 212 IMEDGEISQVLPGDNKFNITCSGYESHVSGGILTSISVAT--PIPGYAYASLALTR 269
 QY 161 PMFR-----TAYVAANVENKEMSLD-----VYNDLI---EWLNLADRRYVDS 203
 DB 270 PVSRLFKNNSILYVFSGNGPKASGDYCIQSNIVFSDEIPASQDMPTVTTDTITVYGDN- 328
 QY 204 QMINFNDDTYVRIKVALPTVDVPTPEGVLTVSDYKLTAYIKCEANMPLVDQGW 263
 DB 329 -----ATYSVPVTSB-----DANSPVTVTAFW 352
 QY 264 -----IGQYALPT----- 273
 DB 353 AMPNNTETDPKCKMTLISGTPSCENISAFASNRTPDITVSGLTAPKTLITRTATNA 412
 QY 274 -----SLPQDVSEAVYALHTLTPAPSSAALAFWAGLPQGTAPAGTPANEQAS 324
 DB 413 TTTTHKVIKFAPESTTSTPTLNTTGFADPNTT-----TGIPSTHVPNTLTA--PAS 463
 QY 325 SGGYLWIRNGTTFPAGSVSYVLPBGFALERVDPNQSW-----TDPASAGDV 373
 DB 464 TGPVSTADVSTPPTAGTSGASP-----VTPSPSPMDNGTGRAPDMTSTSTSVTTP 516

QY 374 TFRQYADVBNVT-----NNPAGGSAFTPTVRVPPSNAYNTYFR 414
 DB 517 TENATSPPAVTTTPNATSPTPAVTTPTPNATSPTLAKTSPTSAVTTPTPNATSPTLGX 576
 QY 415 N-----TLSTRBSRLBLPMPADGGQYANNPKIROSILKELTLCY 458
 DB 577 TSPTSAVTTTPNATSPTLKTSPTS-AVTTPTPAT-GPTVGETSP-QANATNTTLG-- 631
 QY 459 LVHSKRRPVPQLTPASSFGAVSFNNPGYERTRDLPDYTGIRDSFDONNSTAVAHFRSL 518
 DB 632 ---GTSPTPVVTSQPKATSAVTTGQ--HNTSSSTSMGLRPSNPETLSPTSNDNST 686
 QY 519 HSCSIVTITYOGWEGVTNNTPFGQFAHAGLKNREILCLADLATRLTGVYPATDNPA 578
 DB 687 HMPILTSAPHPCGENITGV--TPASISTR-----HVSTSPBRPFGTTSQASGRGNSST 738
 QY 579 AVSAPAAAMLSVLKSEATSSIKSGVETAYGAAS 614
 DB 739 STKPGRVNVTGTPPQKATSPQAPSGQRTAVPTVS 774

RESULT 12
 US-09-328-599A-1

Sequence 1, Application US/09328599A
 ; Patent No. 6432679
 ; GENERAL INFORMATION:
 ; APPLICANT: MOND, James J. and Lees, Andrew
 ; TITLE OF INVENTION: Enhancement of B Cell Activation by
 ; TITLE OF INVENTION: Co-Ligation of Receptors for Antigen and Complement C3d
 ; TITLE OF INVENTION: Using BBV gp350/220 or BBV gp350/220 Peptide Adjuvants
 ; NUMBER OF SEQUENCES: 2
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESS: Finnegan, Henderson, Farbow, Garrett &
 ; STREET: 1300 I Street, N.W., Suite 700
 ; CITY: Washington
 ; STATE: DC
 ; COUNTRY: USA
 ; ZIP: 20005
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/328,599A
 ; FILING DATE:
 ; CLASSIFICATION:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Fordle, Jean B.
 ; REGISTRATION NUMBER: 32,984
 ; REFERENCE/DOCKET NUMBER: 04995.6025-00000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)408-4000
 ; TELEFAX: (202)408-4400
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 907 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-328-599A-1

Query Match 3.5%; Score 119.5; DB 4; Length 907;
 Best Local Similarity 19.5%; Pred. No. 0.049;
 Matches 136; Conservative 67; Mismatches 274; Indels 219; Gaps 27;
 QY 48 NFTAADLAQSLDANTVTFFPNNISMPFRMAKIDLDSDISIKYFKYIDPAGATES 107
 DB 169 NITAVVR--AAGLD--VTLPUSLPTSADSNFS--VKTEMLD--NEID---IEC 211


```

; APPLICANT: Udagawa, Hiroaki
; APPLICANT: Frandsen, Torden
; APPLICANT: Nielsen, Tom
; APPLICANT: Kaupinen, Markus
; APPLICANT: Christensen, Soeren
; TITLE OF INVENTION: Lyso-phospholipase
; FILE REFERENCE: 5958.210-US
; CURRENT APPLICATION NUMBER: US/09/687.5388
; CURRENT FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: Patent version 3.1
; SEQ ID NO 8
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
; US-09-687-538B-8

Query Match      3.4%; Score 116; DB 4; Length 641;
Best Local Similarity 19.3%; Pred. No. 0.057;
Matches 146; Conservative 77; Mismatches 233; Indels 300; Gaps 39;

4 AGV-ASGRPHNRGRTR-----NVRSANVTYVNGRRN----- 34
15 SGVTAAPGPHGERIRIRIDRVLERALPNAPDGYVSNVSCANRPTVSASSGSLNETS 74
35 --QRRRTGROVSPDNF---TAAODLAQSLDANTVTTPANISSMPEFRNMAKKIDIDS 89
75 WLKTRREKTOSAMKDFNHNVTIKDFDAVOYLDNHS-----SNTSNLPNIGIAVSG----- 124
90 DSIQGYFYKLDPAQ-----TESARAVGEYSKIPDGLVKSVDAB-----IRRIY 134
125 ---GQYRLMNGAGAIKAFDSRTENSTNTGQLG---GLDSATYTLAIGSGGMLVGSIT 177
135 NEECPVTVDSVPLDGRQMSL--SIFSP-----MFRATYAAVANVEN----- 175
178 INNFTTIALQTHEBGAVWQFQNSIFEGPDGSIQILDSATYTKHYDAVQDKXAGYET 237
176 -----KEMSLDVN-----DLIEMLN-----NLADWRY----- 198
238 SITDYGWALSYQLINATDGGPSYTWSSIALDTFKQADMPLLVADGRYPDELVSNN 297
199 --VVDSQEWINFNTDITY--VRIYRLRPTVD--VPDTEGLVTRVSDYRLTYKATCE 251
298 ATVEFENMERGETPPTIYGFVPLEYVSGSKFDGGSIPD-NETCVAGFD----- 344
252 ANMPTLVDOGFNIGGOVAL-----TPTSLPOYDVSEAYALHTLTPAPSSAALAFV 303
345 -----NAGFVWGTSSSLFNQFLQVNSTSLPDF-LKTAFSDLAKIGEEDBIA--V 393
304 WAGLFQGGTAPAGTAPMEQAS---SGG-----YLTWRHNGTTPPAGS---VSY 345
394 YAPNFPYMAFVSPDAHQBELMDVYDGEDLQNIPLHPLIQPERHVDVFAVDSADITY 453
346 VLPEGFALERYDPNNGSWTDFASAGDTVFRQVAVDEVVVTNNPAGGSGAPFTYRVPS 405
454 SWEPTAL-----VATY-----ERSLNSGTGANGISFPA---IFQ 486
406 NAVTTFVFRNTLLETPSSRRLELPMPPADFCQTAVANNPKIEOSLKETIAGCYLVSKR 465
487 NTFVANG-----LNRPT-----PFGCNSNTTGP 511
466 NPFQULTPASSFGAVSPNN---PGYE-RYTRDLPDYTGIRDSF-----DONNST 509
512 TPLVVYLLP--NYPYVSYSNMSTFOPSYEISERD---DTRNGYDVVTWNGSTRDQMTT 565
510 AAHFRSLSHSGSIVTKYQG-----WEGVTNVTPPGQFAHAGLKN 552
566 CV-----GCALISRSFERTNTQVPDQCTQCFQYKTCMDGTTNSNP-----ADY 608
553 EELCLADDLATRLTGVPATDNFAAFAVAFANML 588
609 EPTVLLBEDSAGSALS---PAVITTVATSAALFTLL 641
```

```

RESULT 15
US-09-268-347-30
; Sequence 30, Application US/09268347
; Patent No. 6335182
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena M.
; TITLE OF INVENTION: RECOMBINANT HAEMOPHILUS INFLUENZAE ADHESIN PROTEINS
; FILE REFERENCE: 1038-860
; CURRENT APPLICATION NUMBER: US/09/268.347
; CURRENT FILING DATE: 1999-03-16
; NUMBER OF SEQ ID NOS: 54
; SOFTWARE: Patent Ver. 2.0
; SEQ ID NO 30
; LENGTH: 1004
; TYPE: PRT
; ORGANISM: Haemophilus influenzae
; US-09-268-347-30

Query Match      3.4%; Score 116; DB 4; Length 1004;
Best Local Similarity 19.5%; Pred. No. 0.12;
Matches 135; Conservative 92; Mismatches 271; Indels 194; Gaps 31;
```

```

30 NGRNRORRRTRGROVSPDNFTAAOD--LAQSLDAN---TTPP-----ANISMPER 77
129 NGTRNEK---SQYKHADEVLPBGKQGVTVTSKSENGKHVTFLEKDLNFKATVSDKL 185
78 RNWAKG-KIDLDSDISGYFYKLDPAQATSBARAVGEYSKIPD---GLVKSVDARIKEI 133
186 SLGANGNKVDTISDITNGLKFAKPSFNGQNGVHLNGIASLTLDITITGTTKATNGVDQN 245
134 YNECPVTVDSVPLDGRQMSLSPFPMTATYAAVANVENKEMSLDVN--DLIEMLN 191
246 HNRASVADVANA-----GW-----NIGNGASVDVFVNTYDVFVN 282
192 NLADWRYVDSQEWINFNTDITY---YRIYRLRPTVDVDPDTEGLVTRVSDYRLTYK 246
283 GL-----NTNVNTTTPAHNKKTIVKDVGLPQYVTEBGETVYKAGNEY---YE 330
247 A-ITCEANMPTLVDOGFNIGGOVALTPTSL-----POYDVSEAYALH 287
331 AKQDSADMCKVY-----NGKLAKTVKVLVANGNTNPKLSNADGTEMTDAVSFKQLK 385
288 TLTPAPSSAALAFWAGLFQGGTAPAGTAPMEQASGGYLTWRHNGTTPPAGSVSYVL 347
386 ALQDKQVTLASNNYA-----NGGS-----DADQKGIQTLISGLNP----- 422
348 PEGFALERYDPNNGSWTDFASAGDTVFRQ-----VADE 382
423 -----KFKSTDEGLNITKABNDITVTPPKKGSYQVDDGKATIQDQAKTTGLVEASE 475
383 VVVTNNPAG-----GGSAPFTYRVPSNAYNTVFRNTLLETPSSRRLELPMPPAD 435
476 LVDLSLTKGMKVYGTGDTGVT-----DGHHTDLTVKSGDKVTLKAGDNLKVQBGNT 528
436 FQGTAAAN-PKIBOSLKETL-GCYLVHSKRNPFQULTPASSFGAVSPN----- 483
529 FTYALAKDLTVKVSVEFDQDTANANGASTKITKQGLITTPNAGAAAGANANTISVTKD 588
484 --NPGYERTDLPDVTGIRDSFPDQNMSTAVAHFRLSHSGSIVTKY-QGMEGVTVNN-- 538
589 GISAGNKAIVKV--VSLAKTGED-----ANFDELTSADNLTQYDNATYKGLTNLBK 639
539 -----TPP---GQFAHAGLKNBEILCLADDLATRLTGVPATDNFAAFAVAFANMLSS 590
640 SKGQCTPTVADNTATVDDLGIGWVISADTKYKGLNRYNAQVANAIVKSKSGANGINV 699
591 VLKS-----EATSSIIKSVGETAVGAQSG 616
700 SGTILNCGTRITTEFLADBNALIAFSGSKAL 731
```

Fri Jan 16 08:13:26 2004

us-09-991-262-50.rat

Page 11

Job time : 14.0081 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using SW model

Run on: January 15, 2004, 16:52:13 ; Search time 77.607 Seconds
(Without alignments)
1704.735 Million cell updates/sec

Title: US-09-991-262-50

Perfect score: 3374
Sequence: 1 MGDAGVASQPHNRGTRNV.....GKIARVRRARRARRARAN 647

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications AA:*

```

1: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09A_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubppaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubppaa/US10C_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3374	100.0	647	11	US-09-991-262-50
2	3033	89.9	675	11	US-09-991-262-52
3	121.5	3.6	889	12	US-10-174-677-42
4	121.5	3.6	889	12	US-10-295-027-1229
5	121.5	3.6	2344	9	US-09-815-242-12713
6	117.5	3.5	609	12	US-10-369-493-22806
7	117.5	3.5	609	12	US-10-369-493-22806
8	117.5	3.5	26926	10	US-09-759-5088-2
9	117	3.5	860	12	US-10-369-493-18823
10	116.5	3.5	781	12	US-10-187-253A-33
11	116	3.4	641	15	US-10-309-437-8
12	113.5	3.4	2283	12	US-10-172-502-4
13	112.5	3.3	503	10	US-09-738-626-6045
14	112.5	3.3	1443	15	US-10-245-802-18
15	111.5	3.3	781	12	US-10-187-253A-27

16	111	3.3	20	11	US-09-991-262-24	Sequence 24, Appl
17	111	3.3	541	15	US-10-230-026-10	Sequence 10, Appl
18	111	3.3	620	12	US-09-788-051-7	Sequence 7, Appl
19	111	3.3	636	12	US-09-788-051-4	Sequence 4, Appl
20	110.5	3.3	1300	9	US-09-815-242-4903	Sequence 4903, Ap
21	110.5	3.3	1300	9	US-09-815-242-10906	Sequence 10906, A
22	109.5	3.2	1082	12	US-10-369-493-17890	Sequence 17890, A
23	109	3.2	780	12	US-10-289-762-920	Sequence 920, App
24	109	3.2	1250	15	US-10-156-761-7572	Sequence 7572, Ap
25	109	3.2	1463	10	US-09-971-536-69	Sequence 69, Appl
26	108	3.2	591	11	US-09-821-616-7	Sequence 7, Appl
27	108	3.2	1260	15	US-10-245-802-8	Sequence 8, Appl
28	107.5	3.2	532	15	US-10-156-761-11686	Sequence 11686, A
29	107	3.2	690	12	US-10-262-083-2	Sequence 2, Appl
30	107	3.2	690	12	US-10-262-083-18	Sequence 18, Appl
31	107	3.2	2297	15	US-10-245-802-20	Sequence 20, Appl
32	106.5	3.2	478	12	US-10-238-075-657	Sequence 657, App
33	106.5	3.2	1806	15	US-10-156-761-13509	Sequence 13509, A
34	106	3.1	618	11	US-09-821-616-34	Sequence 34, Appl
35	106	3.1	1138	12	US-10-021-660-101	Sequence 101, App
36	106	3.1	1138	12	US-10-394-322A-63	Sequence 63, Appl
37	105.5	3.1	631	12	US-10-369-493-10154	Sequence 10154, A
38	105.5	3.1	1079	15	US-10-112-488-39	Sequence 39, Appl
39	104.5	3.1	1481	12	US-10-050-763-1	Sequence 1, Appl
40	104.5	3.1	1291	12	US-10-366-547-83	Sequence 83, Appl
41	103.5	3.1	1649	12	US-10-369-493-18460	Sequence 18460, A
42	103.5	3.1	2364	15	US-10-156-761-7834	Sequence 7834, Ap
43	103.5	3.1	5245	12	US-10-329-079-45	Sequence 45, Appl
44	103	3.1	2200	12	US-09-796-575-2	Sequence 2, Appl
45	102	3.0	984	15	US-10-156-761-13822	Sequence 13822, A

ALIGNMENTS

```

RESULT 1
US-09-991-262-50
Sequence 50, Application US/09991262
Publicatlon No. US20030041349A1
GENERAL INFORMATION:
APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
TITLE OF INVENTION: Insect Viruses and Their Uses in
Protecting Plants
NUMBER OF SEQUENCES: 57
CORRESPONDENCE ADDRESS:
ADDRESSER: Flehr Hobbach Test Albritton & Herbert LLP
STREET: Four Embarcadero Center, Suite 3400
CITY: San Francisco
STATE: California
COUNTRY: United States
ZIP: 94111-4187
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentln Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/991,262
FILING DATE: 20-NO. US20030041349A1-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/234,238
FILING DATE: 20-JAN-1999
APPLICATION NUMBER: US 08/485,355
FILING DATE: 07-JUN-1995
APPLICATION NUMBER: US 08/440,522
FILING DATE: 12-MAY-1995
APPLICATION NUMBER: US 08/089,372
FILING DATE: 08-JUL-1993
APPLICATION NUMBER: AU PL4081/92
FILING DATE: 14-AUG-1992
ATTORNEY/AGENT INFORMATION:
NAME: Treacartin, Richard F.

```

REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-58631-3/RFT/DSS
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TELE: 910 277299
 INFORMATION FOR SEQ ID NO: 50:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 647 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 50:
 US-09-991-262-50

Query Match 100.0%; Score 3374; DB 11; Length 647;
 Best Local Similarity 100.0%; Pred. No. 1,4e-301;
 Matches 647; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MGDAGVASQRPNNRRGTRNVRVSANTVTVNGRRNRRTGRQVSPEDNFTAAADLAQSL 60
 DB 1 MGDAGVASQRPNNRRGTRNVRVSANTVTVNGRRNRRTGRQVSPEDNFTAAADLAQSL 60
 QY 61 DANTYTFPANISSMPEFRWAKGKIDLDSDSIGWYFKYLDPAQATSAVAAGEYSKIPDG 120
 DB 61 DANTYTFPANISSMPEFRWAKGKIDLDSDSIGWYFKYLDPAQATSAVAAGEYSKIPDG 120
 QY 121 LKFSVDAREIREINEECVNTDVSPLDGRQMSISFSPMERTAYAVANVENKMSL 180
 DB 121 LKFSVDAREIREINEECVNTDVSPLDGRQMSISFSPMERTAYAVANVENKMSL 180
 QY 181 DVVNDLIEMLNLAADRWYVDSBQWINTNDTTYVRIRVLRPTVDVDPDTEGLVRTVSD 240
 DB 181 DVVNDLIEMLNLAADRWYVDSBQWINTNDTTYVRIRVLRPTVDVDPDTEGLVRTVSD 240
 QY 241 YRLTYKAITCEANMPTLVDOGFWMIGQVALTPISLPQYDVSEAYALHTLTFARPSSAAL 300
 DB 241 YRLTYKAITCEANMPTLVDOGFWMIGQVALTPISLPQYDVSEAYALHTLTFARPSSAAL 300
 QY 301 AFVMAGLPQGGTAPAGTAPAMEQASGGYLTWRHNGTTPAGSVSYLPEGFALERDPPND 360
 DB 301 AFVMAGLPQGGTAPAGTAPAMEQASGGYLTWRHNGTTPAGSVSYLPEGFALERDPPND 360
 QY 361 GSWTDFASAGDTVTFROVAVDEVVYTNPNAGGSGAPFTFRVPPSNAYNTVFRNTLLET 420
 DB 361 GSWTDFASAGDTVTFROVAVDEVVYTNPNAGGSGAPFTFRVPPSNAYNTVFRNTLLET 420
 QY 421 RSSRRLLELPMPPADGGQTVANNPKIEQSLKETLCTLVHSMRNPVQLTPASSFGAV 480
 DB 421 RSSRRLLELPMPPADGGQTVANNPKIEQSLKETLCTLVHSMRNPVQLTPASSFGAV 480
 QY 481 SFNNPEYERTDLPDVTGIRDSFDQMSSTVAHFRSLHSCSLVTITYGMEGVTVNVP 540
 DB 481 SFNNPEYERTDLPDVTGIRDSFDQMSSTVAHFRSLHSCSLVTITYGMEGVTVNVP 540
 QY 541 FCGFAHAGLLKNEBELCLADDLATRLTGVPATDNFAAUSAFAANMLSSVLKSEATSS1 600
 DB 541 FCGFAHAGLLKNEBELCLADDLATRLTGVPATDNFAAUSAFAANMLSSVLKSEATSS1 600
 QY 601 IKSVEGTAVGAAGSLAKLPGLLMSVPGKIAARVARRARRARRAARAN 647
 DB 601 IKSVEGTAVGAAGSLAKLPGLLMSVPGKIAARVARRARRARRAARAN 647

RESULT 2
 US-09-991-262-52
 Sequence 52, Application US/09991262
 Publication No. US20030041349A1
 GENERAL INFORMATION:
 APPLICANT: Christian, P. D., Gordon, K. H.J., Hanzlik, T. N.
 TITLE OF INVENTION: Insect Viruses and Their Uses in
 Protecting Plants
 NUMBER OF SEQUENCES: 57

CORRESPONDENCE ADDRESS:

ADDRESSEE: Flehr Hobach Test Albritton & Herbert LLP
 STREET: Four Embarcadero Center, Suite 3400
 CITY: San Francisco
 STATE: California
 COUNTRY: United States
 ZIP: 94111-4187

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/991,262
 FILING DATE: 20-NO. US20030041349A1-2001
 CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/234,238
 FILING DATE: 20-JAN-1999
 APPLICATION NUMBER: US 08/485,355
 FILING DATE: 07-JUN-1995
 APPLICATION NUMBER: US 08/440,522
 FILING DATE: 12-MAY-1995
 APPLICATION NUMBER: US 08/089,372
 FILING DATE: 08-JUL-1993
 APPLICATION NUMBER: AU PL4081/92
 FILING DATE: 14-AUG-1992
 ATTORNEY/AGENT INFORMATION:
 NAME: Treacartin, Richard F.
 REGISTRATION NUMBER: 31,801
 REFERENCE/DOCKET NUMBER: A-58631-3/RFT/DSS

TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 781-1989
 TELEFAX: (415) 398-3249
 TEL: 910 277299

INFORMATION FOR SEQ ID NO: 52:

SEQUENCE CHARACTERISTICS:

LENGTH: 675 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 52:

US-09-991-262-52

Query Match 89.9%; Score 3033; DB 11; Length 675;
 Best Local Similarity 100.0%; Pred. No. 3.8e-270;
 Matches 580; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 68 PANISMPBFNNWAKGKIDLDSDSIGWYFKYLDPAQATSAVAAGEYSKIPDGKLVKFSVD 127
 DB 68 PANISMPBFNNWAKGKIDLDSDSIGWYFKYLDPAQATSAVAAGEYSKIPDGKLVKFSVD 127
 QY 96 PANISMPBFNNWAKGKIDLDSDSIGWYFKYLDPAQATSAVAAGEYSKIPDGKLVKFSVD 155
 DB 96 PANISMPBFNNWAKGKIDLDSDSIGWYFKYLDPAQATSAVAAGEYSKIPDGKLVKFSVD 155
 QY 128 AIREIRINEECVNTDVSPLDGRQMSISFSPMERTAYAVANVENKMSLDVNDLI 187
 DB 128 AIREIRINEECVNTDVSPLDGRQMSISFSPMERTAYAVANVENKMSLDVNDLI 187
 QY 156 AIREIRINEECVNTDVSPLDGRQMSISFSPMERTAYAVANVENKMSLDVNDLI 215
 DB 156 AIREIRINEECVNTDVSPLDGRQMSISFSPMERTAYAVANVENKMSLDVNDLI 215
 QY 188 EWLNNLAADRWYVDSBQWINTNDTTYVRIRVLRPTVDVDPDTEGLVRTVSDVRLTYKA 247
 DB 188 EWLNNLAADRWYVDSBQWINTNDTTYVRIRVLRPTVDVDPDTEGLVRTVSDVRLTYKA 247
 QY 216 EWLNNLAADRWYVDSBQWINTNDTTYVRIRVLRPTVDVDPDTEGLVRTVSDVRLTYKA 275
 DB 216 EWLNNLAADRWYVDSBQWINTNDTTYVRIRVLRPTVDVDPDTEGLVRTVSDVRLTYKA 275
 QY 248 ITCEANMPTLVDOGFWMIGQVALTPISLPQYDVSEAYALHTLTFARPSSAALAFVMAGL 307
 DB 248 ITCEANMPTLVDOGFWMIGQVALTPISLPQYDVSEAYALHTLTFARPSSAALAFVMAGL 307
 QY 276 ITCEANMPTLVDOGFWMIGQVALTPISLPQYDVSEAYALHTLTFARPSSAALAFVMAGL 335
 DB 276 ITCEANMPTLVDOGFWMIGQVALTPISLPQYDVSEAYALHTLTFARPSSAALAFVMAGL 335
 QY 308 PGGTAPAGTAPAMEQASGGYLTWRHNGTTPAGSVSYLPEGFALERDPPNDGSWTDF 367
 DB 308 PGGTAPAGTAPAMEQASGGYLTWRHNGTTPAGSVSYLPEGFALERDPPNDGSWTDF 367
 QY 336 PGGTAPAGTAPAMEQASGGYLTWRHNGTTPAGSVSYLPEGFALERDPPNDGSWTDF 395
 DB 336 PGGTAPAGTAPAMEQASGGYLTWRHNGTTPAGSVSYLPEGFALERDPPNDGSWTDF 395
 QY 368 SAGDVTTFROVAVDEVVYTNPNAGGSGAPFTFRVPPSNAYNTVFRNTLLETSSRRL 427
 DB 368 SAGDVTTFROVAVDEVVYTNPNAGGSGAPFTFRVPPSNAYNTVFRNTLLETSSRRL 427
 QY 396 SAGDVTTFROVAVDEVVYTNPNAGGSGAPFTFRVPPSNAYNTVFRNTLLETSSRRL 455
 DB 396 SAGDVTTFROVAVDEVVYTNPNAGGSGAPFTFRVPPSNAYNTVFRNTLLETSSRRL 455
 QY 428 ELPMPADPGQTVANNPKIEQSLKETLCTLVHSMRNPVQLTPASSFGAVSFNNPEY 487
 DB 428 ELPMPADPGQTVANNPKIEQSLKETLCTLVHSMRNPVQLTPASSFGAVSFNNPEY 487

```

Db      456 ELMPADDEGQYANNPKIQSLKELTLCYLVHSHKRNVPQUTLPPSSGCAVSFNNPGY 515
Qy      488 ERTDRLPDYTGIRDSFDONNSTAVAHFRSLSHSCSIYTKYQMEGVNTNTPFGQPAHA 547
Db      516 ERTDRLPDYTGIRDSFDONNSTAVAHFRSLSHSCSIYTKYQMEGVNTNTPFGQPAHA 575
Qy      548 GLKNEEIIICLADDLATRLTGVPATNPFANAASAFANMLSVLKSEATSSIIKSGET 607
Db      576 GLKNEEIIICLADDLATRLTGVPATNPFANAASAFANMLSVLKSEATSSIIKSGET 635
Qy      608 AVGAAGSLAKLPGLMSVPEKTIARVARRARRARRAN 647
Db      636 AVGAAGSLAKLPGLMSVPEKTIARVARRARRARRAN 675

```

RESULT 3

```

US-10-174-677-42
; Sequence 42, Application US/10174677
; Publication No. US20030190704A1
; GENERAL INFORMATION:
; APPLICANT: Xie, Ting
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR ANCHORING STEM CELLS IN A MICROENVIR
; FILE REFERENCE: 40716(IP-012)
; CURRENT APPLICATION NUMBER: US/10/174,677
; CURRENT FILING DATE: 2002-06-19
; NUMBER OF SEQ ID NOS: 117
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 42
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-174-677-42

```

```

Query Match      3.6%; Score 121.5; DB 12; Length 889;
Best Local Similarity 20.6%; Pred. No. 0.099;
Matches 145; Conservative 91; Mismatches 250; Indels 219; Gaps 36;

```

```

Qy      32 RRNRRTGRQVSPDNFTAAQDLASLDANTVTPANISSMPEFRNWKAG----- 83
Db      97 RHNAKCOLSLSEVFANDKEIKIKVEIQDINDNAPSSDQIEMDISRNAAGTRFPLTSA 156
Qy      84 -KIDLDSDSIGMYKYVLDPAAGATE-SARAVGEYSKIPDGLVKEFVDAEIRIYNEBCPV 141
Db      157 HDPDAGENGLRTYLLTRDDHGLFGLDVKSQDGTKEFELVYQKALD---REQQHHTLV 213
Qy      142 TDVSVPLDGRQWSLISFSPMFTAYAVANVENKEMS--LDVNNDLIEMLNLAIDRWYV 199
Db      214 T-----ALDGE-----PPRSATVOINVKYIDSDNSPVREASVYELPENAPLGTVV 262
Qy      200 VDSEQWINTN-----DTTY-----YVRIRVLAPTYDVPDPTGELVTVS--DY----- 241
Db      263 ID-----LNATDADGPNGEVLXSFSSVYPRV-RELPSI-DPKGLIRKGNLDYBEENG 316
Qy      242 -----RTTYAITCEAMPFLVNOGFITGQYALTPISLQYDVE 282
Db      317 LEIDVQARDIGPNPIPAHCKVTYKLIIDRNDAFPI--GF-----VSVQGALSE 363
Qy      283 AYALHTLTFAAPSSAALAFVW---AG-----LPQGTAPAGTPAMEQASSGYLT 330
Db      364 A-----APRGIVIALVRYTDRDSDGNGLQCRVLGGSGTGGG-----VSTVQALSE 401
Qy      331 WRHNGTTPAGSVSYVLPEGFALERYDPNDGWTDFASAGDTVTFROY--AVDEVVNTN 387
Db      402 -----GLGGPGGSVPFKLEENY-----DNFYT-----VYIDRPLDRETQDEYNTVI 442
Qy      388 NPAGGSAP-----TFYRV-----PR--SNAVTNTVFNTILTRPSSRRLELPMRP 433
Db      443 VARDGSPPLNSTYSFALKIILDENDNPRFTKGLVYLQVHENNI-----PG 488
Qy      434 ADPGQYANNPKIRQ-----SLKETLGLYLVHSHKRN-NPQFOLTPASSFGAVSFNNPG 486
Db      489 EYLGSLVAQDBDLQNGTVSYSLPSHIGDVSITYYSVNP-----TNKAITAALSFN--- 541

```

```

Qy      487 YERTDRLPDYTGIRDS-----FDONNSTAVAHFRSLSHSCSIYTKYQMEG----- 533
Db      542 FEQTKAFBEFKVLAKDSGAPAHLESNATVTLVDNDNAVITLPTLQNDTAELQVPRNA 601
Qy      534 -----VTNVTPPGQPAHAGILK-----NEEIIICLADDLATRLTGVPATDN-----P 576
Db      602 GLGLVSTVRAALSDSGESRLTYEIVDGNDDLPIIDPSGSGIRTLHPFMEVDVTVPEL 661
Qy      577 AAIVSAFANMLSSVLSKERTSSIIKSGEATVGAAGSLAKLP 621
Db      662 VKATDHGKPTLSAVAK-----LIIRSVS-----GSLPEGVPRVNG 697

```

RESULT 4

```

US-10-295-027-1229
; Sequence 1229, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Nacasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gleh, Kurt C.
; APPLICANT: Glyme, Richard
; APPLICANT: Hevezl, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Bos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of diagnosis of Cancer, Compositions and
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PAM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1229
; LENGTH: 889
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1229

```

```

Query Match      3.6%; Score 121.5; DB 12; Length 889;
Best Local Similarity 20.6%; Pred. No. 0.099;
Matches 145; Conservative 91; Mismatches 250; Indels 219; Gaps 36;

```

```

Qy      32 RRNRRTGRQVSPDNFTAAQDLASLDANTVTPANISSMPEFRNWKAG----- 83
Db      97 RHNAKCOLSLSEVFANDKEIKIKVEIQDINDNAPSSDQIEMDISRNAAGTRFPLTSA 156
Qy      84 -KIDLDSDSIGMYKYVLDPAAGATE-SARAVGEYSKIPDGLVKEFVDAEIRIYNEBCPV 141

```

```

Db      157 HDPDAGENGLRTYLLTRDDHGLFGLDVKSRGDTKPELVIOKALD---REOQNHHTLV 213
Qy      142 TVSVPLDGRWMSLISFSPMERTAYAAVANKEMS--LDVNDLIMLNMADMRVY 199
Db      214 T-----ALDGE-----PPRSATVQINKVIDSNNSVPEKAPSLVELPENAPLGTV 262
Qy      200 VDEQWMINFTN-----DTTY-----YAIRVLRPTVDVDPTEGLVTVS--DY----- 241
Db      263 ID-----LNATDADDEGNGEVLTFSSSYVDRV-RELFSI-DPKTGLIRVGNLDYRENGM 316
Qy      242 -----RLTKATTCENMPTLVDOGFWIGQVALTPISLPQDVSE 282
Db      317 LEIDVQARDLGNPIPAHCKVYKILDRDNMAESI---GF-----VSVRQALSB 363
Qy      283 AVALHTLPARPSAALAFVW---AG-----LPQGTAPAGTPAWEQASGGYLT 310
Db      364 A-----APPTVIALVRVTRDRSGKXGQLQCRVLGGGTGGG----- 401
Qy      331 WRHNGTTPAGSVSYVLPBGFALERVDNDGWTDPASAGDTVTFQV--AVDEVVTN 387
Db      402 -----GLGPGGSVPFKLENY-----DNFYT-----VVTDRPLDETQDEVNTII 442
Qy      388 NAGGGSAP-----TTTVRV-----PP---SNAVTNTVPRNTLLETRESRLLEMPR 433
Db      443 VARDDGSPPLNSTKSFALIKILDENDNPPRFKGLVYLQVHENNI-----PG 488
Qy      434 ADFGQVANNPKIEQ-----SLKKEITGCVLVHSKMR-NPVQOLTTPASSFGAVSPNNQ 486
Db      489 EYIGSVLADDPDLGQNGTYSILPSHIGDVSITYTVSNP-----TNGAIVALARSTN--- 541
Qy      487 YERTRLDPYTGIRDS-----PDQNMSTAVAHFRSLSHSGSIVTKYQWEG----- 533
Db      542 FEQTAKFEFKVLAKDQAPAHLESNATVTVLVDVNDNAVILPTLQNDTALQVPRNA 601
Qy      534 -----VTNVTTPFGQFAHAGLAK-----NEELICLADDLATLGTVPATDN-----F 576
Db      602 GLGYLSTVRALDSDGESGRLYEIVDGNDDLFEIDPSSGBIRTLHPMEDVTPVEL 661
Qy      577 AAASVAFANMLSSVLSKSEATSSIIKSVGETAVGAQSGGLAKPG 621
Db      662 VKNVTHGKRTLSAVAK-----LIIRSVS-----GSLPEGVPRVNG 697

```

RESULT 5
US-09-815-242-12713
Sequence 12713, Application US/09815242
Patent No. US20020061569A1
GENERAL INFORMATION:
APPLICANT: Haelebeck, Robert
APPLICANT: Ohlsen, Karl L.
APPLICANT: Zyskind, Judith W.
APPLICANT: Wall, Daniel
APPLICANT: Trawick, John D.
APPLICANT: Carr, Grant J.
APPLICANT: Yamamoto, Robert T.
APPLICANT: Xu, H. Howard
TITLE OF INVENTION: Identification of Essential Genes in
TITLE OF INVENTION: Prokaryotes
FILE REFERENCE: ELITRA.011A
CURRENT APPLICATION NUMBER: US/09/815, 242
CURRENT FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: 60/191, 078
PRIOR FILING DATE: 2000-03-21
PRIOR APPLICATION NUMBER: 60/206, 848
PRIOR FILING DATE: 2000-05-23
PRIOR APPLICATION NUMBER: 60/207, 727
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: 60/242, 578
PRIOR FILING DATE: 2000-10-23
PRIOR APPLICATION NUMBER: 60/253, 625
PRIOR FILING DATE: 2000-11-27
PRIOR APPLICATION NUMBER: 60/257, 931
PRIOR FILING DATE: 2000-12-22

```

; PRIOR APPLICATION NUMBER: 60/269, 308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 12713
; LENGTH: 2344
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12713

Query Match      3 6%, Score 121.5, DB 9, Length 2344,
Best Local Similarity 20.2%, Pred. No. 0.47,
Matches 142, Conservative 81, Mismatches 322, Indels 159, Gaps 31,

Qy      22 VSANTVTVNGRNRORRTRGRQVSPDNFTAA---AQDLAQSLDANTVTPANISSMPEPR 78
Db      254 LFRNTITVKKDLKQYMTTSGNATYDQSGVTVTLQDTTSQGA--ITLGTIRIDSKSFH 311
Qy      79 NNAKGIIDS-----DSIGYFK--YLDPAGATESARAVGEYSKIPDGLVKFSVD 127
Db      312 --FSGKVNIGNKYEGHNGDGDIGFAPSGVLGETGLNGAAVIGGLS---NAFGKLD 365
Qy      128 AIRRIYNEEC-----VTVDSVPLDGRWMSLISFSPMERT-----AYVANVTE 174
Db      366 T-----YHNTSTNNSAKAKADPSNVAQGG-----AFAPVTTDSYGVASTYSSSTAD 414
Qy      175 N-KENSLDVNDLIE-----MLNLMADMRVYVDSQWIMF----- 208
Db      415 NNAKLNQVQTNTPQDFDINTNGDKMTVKYAGQWTFNIDW---IAKSGTNPSSLM 471
Qy      209 ---TNDTYVIRVLRPTVDVDPTEGLVTFVSRLTYKAITCEANMPTLVDOGFWIG 265
Db      472 TASTGATNLQGVQF--GTFEYTESAVTQVRYVD--VTGKDIIPKTVSGNVDQVVTID 527
Qy      266 GQY-ALTPTSLQVOYSEYAL-----HTLPARPSAALAFVWAGLPQGTAPAGTP 318
Db      528 NQOSALTANQYNTSVDSSTYASTNDNTKTKMTNAGQSVTTFYFDVKAP---YTVVGNQ 584
Qy      319 AWEQASGG--YLWRHNGTTPAGSVSYVLPBGFALERYPDNDG----- 362
Db      585 TIEVGKTNPIVLTITDNQGTGVNTVT-GLISGLS---YDARTNIIIGTPKIGSYVT 640
Qy      363 --WTFPASAGDTVTFQVAVDEVVTNNPAGGSAFTTVRV-----DGNATNTV 412
Db      641 VVSTDQANKSTTFTINVVDTTAPVTPIGDKSSSEVFSPISPINIATQDNGNAVTNTV 700
Qy      413 FRNTLLETTPSSRRLLEPMPADPGQTVANNPKIEBSLKERITGCVLVHSKMRNPVQOLT 472
Db      701 -----TGLPSGLTFD-----STNNTTISGTP-----TNIGTSTITTV 731
Qy      473 PASSFGAVFPNNGYERTD-LPDYTGIRDSFPQNMST-----AAHFRSLSHSGSIVTKT 527
Db      732 STDASGNKTTTKYKVTYTNMSMDSVSTSGTQOQSGSVSTRADSGASSTISGSLMTST 791
Qy      528 YQWEGVTVNTPFGQFAHAGLKNKEELICLADDLATRLRGVYPATDPAVAASVAPANN 567
Db      792 SASSTKSTSVLSDSVSAGKS--LSTSESNVSSTSTSVNSQSVSSMSGSGVS--KSTS 848
Qy      588 LGSVLKSEATSSIIKSVGETAVGAQSGGLAKPGLMSVPGKIA 631
Db      849 LSPFISNSSSTKSSBSVSTSDSLKTSLSLSDSVMSYSGSL 892

```

RESULT 6
US-10-369-493-22725
Sequence 22725, Application US/10369493
Publication No. US20030233675A1
GENERAL INFORMATION:
APPLICANT: Cao, Yongwei
APPLICANT: Hinkle, Gregory J.
APPLICANT: Slater, Steven C.
APPLICANT: Goldman, Barry S.
APPLICANT: Chen, Xianfeng

;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
;; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

;; FILE REFERENCE: 38-10(52052)B
;; CURRENT APPLICATION NUMBER: US/10/369,493

;; CURRENT FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039

;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374

;; SEQ ID NO 22725
;; LENGTH: 609

;; TYPE: PRT
;; ORGANISM: Schizosaccharomyces pombe

US-10-369-493-22725

Query Match 3.5%; Score 117.5; DB 12; Length 609;
Best Local Similarity 22.2%; Pred. No. 0.13;
Matches 94; Conservative 47; Mismatches 152; Indels 131; Gaps 23;

QY 224 TYDVPDPTEGLVRY-----SDYRLTYKALICEANMPLVDQGFWIGGQYALTPISLQY 278
DB 138 TVEVVEPLAGVTYTTIHSGSVEYNTT--LATASGTVPGEV-----VEPAGVTYTTIY 190
QY 279 DVSEAYAL-----HTLTFARSSAALAFVWAG-----LPQGTAPAGTPAM 320
DB 191 SGSEFTTTLASAGSISGIVAEIETAGVTYTTISGDOETTLIARSGIVP-GTVEV 249
QY 321 EQASSGGYLTWRHNG-----TFPPA-GSVSYLPEGFALERYDNDSDMTDFASAGDTV 373
DB 250 IEPAGVTYTTIYSGSVEYTTLVDPASGSV-----GTVEVBAVAGVTYTTILQSG--- 300
QY 374 TFRQVANDVAVVNNPAGGSAFTTVAVPSPNATNTVFRN-----TLETRPS 424
DB 301 -----SQAFYTVPASGSVSGTVEVQPTGVTNTVYRSGQTITSTLATASGTVPGT 353
QY 425 RRLLEMPADPGQTVANNPKIEQS-----LKETIGC-----YLW-----SKKN 466
DB 354 VEVLPGSTIYSGTVAITTYDVSTPASTVAVIPTAVCNGERGLQYAVNYIDISSKN 413
QY 467 PVFQLTTPASSFGAV-SFNNPGYERTDL-----PDYTGIRDSFDQ-----NM----- 507
DB 414 ---QCATSGVTDVSSFTQPAVFGSSDLQSSPLFTGVLSSDSVFQMTSSYNLPGYPPD 470
QY 508 STAVAHFRSLHSGSI-----YKTYQ-----GREGYTNV---NTPRG 542
DB 471 ATAMG---STSSACKVIYQFFRVPVDTFSLDVTNVDVDFYGFQKALSGMNTNYD 527
QY 543 QFAH 546
DB 528 TYAY 531

RESULT 7

US-10-369-493-22806

;; Sequence 22806, Application US/10369493

;; Publication NO. US20030233675A1

;; GENERAL INFORMATION:

;; APPLICANT: Cao, Yongwei

;; APPLICANT: Hinkle, Gregory J.

;; APPLICANT: Slater, Steven C.

;; APPLICANT: Goldman, Barry S.

;; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

;; FILE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

;; FILE REFERENCE: 38-10(52052)B
;; CURRENT APPLICATION NUMBER: US/10/369,493

;; CURRENT FILING DATE: 2003-02-28
;; PRIOR APPLICATION NUMBER: US 60/360,039

;; PRIOR FILING DATE: 2002-02-21
;; NUMBER OF SEQ ID NOS: 47374

;; SEQ ID NO 22806
;; LENGTH: 609

;; TYPE: PRT

;; ORGANISM: Schizosaccharomyces pombe

US-10-369-493-22806

Query Match 3.5%; Score 117.5; DB 12; Length 609;
Best Local Similarity 22.2%; Pred. No. 0.13;
Matches 94; Conservative 47; Mismatches 152; Indels 131; Gaps 23;

QY 224 TYDVPDPTEGLVRY-----SDYRLTYKALICEANMPLVDQGFWIGGQYALTPISLQY 278
DB 138 TVEVVEPLAGVTYTTIHSGSVEYNTT--LATASGTVPGEV-----VEPAGVTYTTIY 190
QY 279 DVSEAYAL-----HTLTFARSSAALAFVWAG-----LPQGTAPAGTPAM 320
DB 191 SGSEFTTTLASAGSISGIVAEIETAGVTYTTISGDOETTLIARSGIVP-GTVEV 249
QY 321 EQASSGGYLTWRHNG-----TFPPA-GSVSYLPEGFALERYDNDSDMTDFASAGDTV 373
DB 250 IEPAGVTYTTIYSGSVEYTTLVDPASGSV-----GTVEVBAVAGVTYTTILQSG--- 300
QY 374 TFRQVANDVAVVNNPAGGSAFTTVAVPSPNATNTVFRN-----TLETRPS 424
DB 301 -----SQAFYTVPASGSVSGTVEVQPTGVTNTVYRSGQTITSTLATASGTVPGT 353
QY 425 RRLLEMPADPGQTVANNPKIEQS-----LKETIGC-----YLW-----SKKN 466
DB 354 VEVLPGSTIYSGTVAITTYDVSTPASTVAVIPTAVCNGERGLQYAVNYIDISSKN 413
QY 467 PVFQLTTPASSFGAV-SFNNPGYERTDL-----PDYTGIRDSFDQ-----NM----- 507
DB 414 ---QCATSGVTDVSSFTQPAVFGSSDLQSSPLFTGVLSSDSVFQMTSSYNLPGYPPD 470
QY 508 STAVAHFRSLHSGSI-----YKTYQ-----GREGYTNV---NTPRG 542
DB 471 ATAMG---STSSACKVIYQFFRVPVDTFSLDVTNVDVDFYGFQKALSGMNTNYD 527
QY 543 QFAH 546
DB 528 TYAY 531

RESULT 8

US-09-759-508B-2

;; Sequence 2, Application US/09759508B

;; Publication NO. US20020182599A1

;; GENERAL INFORMATION:

;; APPLICANT: Fishman, Mark C.

;; TITLE OF INVENTION: Methods for Diagnosing and Treating Heart Disease

;; FILE REFERENCE: 00786/381002
;; CURRENT APPLICATION NUMBER: US/09/759,508B.

;; CURRENT FILING DATE: 2001-01-12
;; PRIOR APPLICATION NUMBER: US 60/175,787

;; PRIOR FILING DATE: 2000-01-12
;; NUMBER OF SEQ ID NOS: 11

;; SOFTWARE: PatentIn version 3.1
;; SEQ ID NO 2

;; LENGTH: 26926

;; TYPE: PRT

;; ORGANISM: Homo sapiens

US-09-759-508B-2

Query Match 3.5%; Score 117.5; DB 10; Length 26926;
Best Local Similarity 20.9%; Pred. No. 56;
Matches 125; Conservative 65; Mismatches 218; Indels 191; Gaps 29;

QY 29 VNGRNR-----RTGRQVSPDNFTAAQDLAQGLDANTVTPPANISGMPFRMWA 81
DB 21620 INGNBYQFRVSAVNVKFGVGR---PLDSDPVVAQIOYTPDAPGIPBSNITGNSTLTMA 21676
QY 82 KGIIDLDSDSIGWTFKYLDPAQATESARAVGRYSKIPLGALVPSYDA----- 128
DB 21677 RPESDGSB---IQOYILBRBKSKSTRWVAVISRPISERFKVTGLTBGNEYEFHYMA 21732
QY 129 -----BIREIYNEBCP-----VMTDVSVPLDGRQMSLSIFSPMRTAYV 168


```

Db 21733 ENNAGVBPASGISRLKCEPVPNPPGPTVVKTTDTSKTVSLEMSKPVF----- 21782
Qy 169 AAVANVENKMSLDVNDLLEWLNNLADWERYVDSQEQINFNTDPT-----YYVRIRV 220
Db 21783 -----DOGMELIGYIEMCKTDLDGM-HKVNAEACVTRTYTVDLQAGEYKFRVSA 21833
Qy 221 LRFTYVDPDTE--GLVRYTSDBRLTYKAITCEANMPT--LVDOG---FWIGQYALTP 272
Db 21834 INGA-GKGBSCETVGTIKAVD--RLTAPBLDIDANFKQTHVPAAGSIRLFIAYQGRPTP 21890
Qy 273 T---SLPOYVSEAVYALH-----TLTPAPSSAALAFVWAGLP 308
Db 21891 TAWMSKPDNSLRLADITHDTSFTLVENCNRNDACKTLLTYENNSGSGSITFTYVLD 21950
Qy 309 QGGTAPAGTPAWEQASSG--GYLTWRHNGTTFPAGSVSYVLPEGFALERYDPNDGWSW--- 363
Db 21951 TPG--PPGPIFTPKDVTGSAATLWM--DAPLLDGGARLH---HYVKEKRSRSMQVIS 22002
Qy 364 -----TDPASAGDTYTFROVAVDE-----VVYTNPPAGGSAFTPTVR 401
Db 22003 ECKTRQIFKYNVDLAE--GVPPYFRVSAVNEYGVBEPYEMPEPIVATEQPA----- 22050
Qy 402 VPPSNAYTNTPVFNTELETRPSSRRLLEPMPADFGQTV-----ANNP 444
Db 22051 -PPR-----RUDVDTSKSSAVLWMLKPDHGGSRITGYLLHMRQKSGDLWVEAGHT 22101
Qy 445 K---IEQSLLKETLGCYLVHSKMRNPFQULTPASFCAVSFNNPGYERTDLPDYTGI 499
Db 22102 KOLTFYERLVEKTE--YEFVRVAKNDAGYSEPREASSYLIKPEQLEPATDL---TGI 22155

RESULT 9
US-10-369-493-18823
; Sequence 18823, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 18823
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Anabaena PCC7120
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(860)
; OTHER INFORMATION: unsure at all Xaa locations
US-10-369-493-18823

Query Match 3.5%; Score 117; DB 12; Length 860;
Best Local Similarity 20.9%; Pred. No. 0.24; Indels 192; Gaps 25;
Matches 111; Conservative 54; Mismatches 174;

Qy 233 GLVFTVSDYRLTY-KAITCEANMPTLVDOGFWIGQYALTPSTL-----POY 278
Db 273 GTTMSITDDTLSTYGLSLRVEGNDPT-----GQ-GITPALLEGGINLNGTSNPHI 322
Qy 279 DVSQAY-----AHTLTPAPSSAALAFVWAGLPQO----- 310
Db 323 LVSDAIFGTGDLVSVTAPNGWTAVYTIS--PITDANAAAMTTTPPDLSITVRGFIN 379
Qy 311 GTAPAGTPAWEQASSGGYLTWR-----HNGTTFPAGSVSYVLE--GFALERYD----- 357

```

```

Db 380 NTAIVTSVAPNQTAVNGFTINLAVESPAQTSLSLVANINIAQLPGQTPANNFPYVDSGDQNP 439
Qy 358 -----PDSQWTDPAAGDTVTRQVAVDESVYTNPPAG----- 391
Db 440 SIFNQGQNMTRPALTLANDGVDSLPDTSVDDGFV-DHPAPBETGVGNNNDTVGV 498
Qy 392 GGSAPFTVAVP-----PBN-----YTN----- 410
Db 499 DQGEANITFLVFPVVASAILNGFPVAPNAPNAGPQGGTNDPSSSLVPPTQPNSTLPPQ 558
Qy 411 TVFRNTLETRPSSRRLLEPMPADFGQTVANNPXTLKEITGCTLVHSKQNPV- 468
Db 559 VAFNTIIRSGAALSYSLVPIIPA--SNTBPN--NADLPVGLVITTVASBSYSYWM 613
Qy 469 -----FOL--TPASSFGAVSFN-----NPGYERTDLPDYTGIRDSFPO 505
Db 614 TGSSEFLFDLRNPATLGDOSTINLTSEYITITSVAPGTINNGVEIDLDDGTQLSDYDK 673
Qy 506 NSTAVAHFRSLSHSCSYTKTYQGBWGTNNTPPGQFAHAGLKNBRLICLADLALR 565
Db 674 GEPVPTAP-----VDDGTAGLNGAVQNTTINR--VYGFPLRLRLSLRLDQNGQ 723
Qy 566 LTVGVYDATDNFAAASAPANNMLSVLKEATSSLIK-SVGETAVGAQSG 615
Db 724 VAG-----ADGILSDTDKSPVRGNITIEYQIYNSISEPQSG 759

RESULT 10
US-10-187-253A-33
; Sequence 33, Application US/10187253A
; Publication No. US20030170612A1
; GENERAL INFORMATION:
; APPLICANT: Pichancee, Sergio
; APPLICANT: Snyamala, Venkatakrishna
; TITLE OF INVENTION: DIAGNOSTIC ASSAYS FOR PARVOVIRUS B19
; FILE REFERENCE: CHIR-17194/03US / PPI7194.004
; CURRENT APPLICATION NUMBER: US/10/187,253A
; CURRENT FILING DATE: 2003-03-10
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 781
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: VPI amino acid from
; OTHER INFORMATION: parvovirus B19 clone 2-86
US-10-187-253A-33

Query Match 3.5%; Score 116.5; DB 12; Length 781;
Best Local Similarity 19.3%; Pred. No. 0.23;
Matches 127; Conservative 51; Mismatches 192; Indels 287; Gaps 29;

Qy 112 GYSKIPDGLVKSVDABIR-----EIVNECPVTVDSVPLDGRQMSLSIFSPMPTAY 167
Db 85 GQLSDHPHALSSSSSHAPRPGDAVLSSEDLHCPQGVSVLPQ-----TNY 130
Qy 168 VAAVANENKMSLDVNDLLEWLNNLADWRY-----VDSQEQINFNTDPT 212
Db 131 VGRKN-----ELQAGPQSAVDSAAIRIHDPRYSOLAKGIMPYTHWTVADBELKNINERT 186
Qy 213 TYVVRIRVLRPTVDVDPRTGGLVRYTSDBRLTYKAITCEANMPTLVDOGFWIGQYALTP 272
Db 187 GFAQV-----YKDY--FTLKG-----AAAVVAHFQO-----SL 213
Qy 273 TSPPOYVSEAVYALHTLTPAPSSAALAFVWAGLPQGGTAPAGTPAWEQASSGGYLTWR 332
Db 214 PEPVPAVYASRKY--SNTSVNSAAS-----TGAGGGSNPFYKS--MMSR----- 254
Qy 333 HNGTTFPAGSVSYVLEPGFALERYD----- 358
Db 255 --GATFSANSVTCTFSRQFLIP--YDEPHYKVESPAASGHNASGKAQVCTTISPIMGVS 311

```

```

QY 359 -----NDGSMWDPASAGDTVFROVAVDDEVVYVNNPAGG 393
DB 312 TPRMYLDNMLNLFESPLEFQHLIENYGS---IADPALVTISETIAVKV---TNKYGCG 365
QY 394 -----SAPFTVAV---PPSNAVTNTVFRNLTLET 420
DB 366 VQVDTSTGRICMLVDHEKYPYVLGGQGDTLAPELPIVVFPPQYAVLTIVGDVNT----- 421
QY 421 RPSRRLLEPMPADPGQTVANNPKIBOSLKETTGCTV-LVHSKRNPNVFOULTPASSRGA 479
DB 422 -----QGISGDSK---KLASEEAPFVLEHSS-----FOLLGTCGTAT 456
QY 480 VSFN-----NPGYERTDLDPYTGIRDSFQNMNSTAAHFRSLSH 519
DB 457 MSYKFPVPPEPNEBEGSOHFYEMNPBLGSRIGVPTLG-----GDPKFRSLTH 505
QY 520 -SCSIVTKYQMEGVTVNTPFGQFAHAGLKNSEIICLADDLATRLTGVPYPA----- 572
DB 506 EDHALQPNFMGPPLVNSVSTKEGDSSTGAGKALTGLSTGSTRIS-LAPGVPVSQPY 564
QY 573 ---TDNPAAVSAPFAANMLSSVLKSEATSSIIKSVGETAVGAA-----QSGIANKLP 620
DB 565 HMDTDKVTGGINAI-----SHGQTTYGNAEDKEYQGVGRFP 602

```

RESULT 11

```

US-10-309-437-8
; Sequence 8, Application US/10309437
; Publication No. US20030119164A1
; GENERAL INFORMATION:
; APPLICANT: Udagawa, Hiroaki
; APPLICANT: Frandsen, Torben
; APPLICANT: Nielsen, Tom
; APPLICANT: Kauppinen, Markus
; APPLICANT: Christensen, Soeren
; TITLE OF INVENTION: lysophospholipase
; FILE REFERENCE: 5958.210-US
; CURRENT APPLICATION NUMBER: US/10/309,437
; CURRENT FILING DATE: 2002-12-04
; PRIOR APPLICATION NUMBER: US/09/687,538
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 8
; LENGTH: 641
; TYPE: PRT
; ORGANISM: Aspergillus oryzae
US-10-309-437-8

```

```

Query Match 3.4%; Score 116; DB 15; Length 641;
Best Local Similarity 19.3%; Pred. No. 0.19;
Matches 146; Conservative 77; Mismatches 233; Indels 300; Gaps 39;

```

```

QY 4 AGV-ASORPHNRKGR-----NRAVSANTVYVNGRR----- 34
DB 15 SGVTAAPBPHGERIRIDRTVLERALPNAPDGYVPSNVSCLPNRPVRSAGSLSNETS 74
QY 35 --QRRRTGRQVSPDNF---TAAODLAQSLDANTTTPFANISSMPEFFNMAGKIDLS 89
DB 75 WLKTRREKTSAMKDFFNHVTIKDFDAVOYLDNHS---SNTSNLPNIGIAVSG----- 124
QY 90 DSIGMYFYKLDPA-----TESARAVEYSKIPDGLVKFSVDAB-----IRBIY 134
DB 125 ---GGYRALMNGAGAIKAFDSRTENSTATGQ-----GLQSATYLAGLSGGMLVGSY 177
QY 135 NECCVAVDVSVPLDGRQSL--STFSFP-----MFRATYVAVANVEN----- 175
DB 178 INFTTISALQTHEGAVWQFONSIFEGPDGDSIOILDSATYKVTYAVQDKDAGYET 237
QY 176 ---KEMSLDVVN---DLIEWLN-----MLADWRY----- 198
DB 238 SITDVGALSLYQOLINATDGGPSYTWSSIALIDTRKQADMPLVADDRYDELDVSVSN 297

```

```

QY 199 --VVDSEWINFTNDTTY--VRIRVLAFTYD---VPPRTGLVPTVSDYRLTYKAICE 251
DB 298 ATVHEFNMERGETEPPTYGVFPLETVGSKKDGSSI PD-NETCVAGFD----- 344
QY 252 ANMPTLVQGFWIGGOVAL-----TPSTLPQYDVBAAVALTLTPARPSAALAFV 303
DB 345 -----NAGVFMGSSSLFPQFLQVNSTLPPD-LNTAIBSDILAKIGEBEDTA---V 393
QY 304 WAGLPQGTAPAGTPAMEQAS---SGG-----YLTWRNNGTTPAGS---VSY 345
DB 394 YAPNFPYMAVSPSAHQELMDVGDGEDLQNIPLHPIQBERHVDYFAVDSADTTY 453
QY 346 VLPSFALERDNDGNSWTDPSAGDITYTPQVAVDEVVYVNNPAGGSSAPFTTRVPS 405
DB 454 SMPNGTAL-----VATY-----ERSLNBTGLANGSPPA---IPDQ 486
QY 406 NAYTVTVRNTLLETPSPSRRLLEPMPADPGQTVANNPKIBOSLKETTGCTVHSQR 465
DB 487 NTFVWNG-----LNTFRP-----PFGCNSINTTGP 511
QY 466 NPVFQULTPASSFGAVSFNN---PGYE-RTDLPDYTGIRDSF-----DONNST 509
DB 512 TPLVYYLP--NYPVYSYGNWSTPQSYEISRD---DTIRNGYDVVTMGNSTRDGNMT 565
QY 510 AVAHRSLSHSCSYVTKYQG-----WEGVTNNVTPFGQFAHAGLKN 552
DB 566 CV-----GCALISRSFERTNTOVPDCTQCPQKCMQDGTMTSTNP-----ADY 608
QY 553 BEIICLADDLATRLTGVPATDNFAAVSAPFAANML 588
DB 609 EPVTLIEDSAGSALS---PAVITTVATSAALFTIL 641

```

RESULT 12

```

US-10-172-502-4
; Sequence 4, Application US/10172502
; Publication No. US20030185633A1
; GENERAL INFORMATION:
; APPLICANT: FOSTER, Timothy et al.
; TITLE OF INVENTION: CROSS-REACTIVE MONOCLONAL AND POLYCLONAL ANTIBODIES.
; FILE REFERENCE: P07263US01/BAS
; CURRENT APPLICATION NUMBER: US/10/172,502
; CURRENT FILING DATE: 2002-06-17
; PRIOR APPLICATION NUMBER: US 60/298,098
; PRIOR FILING DATE: 2001-06-15
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 4
; LENGTH: 2283
; TYPE: PRT
; ORGANISM: Staphylococcus epidermidis
US-10-172-502-4

```

```

Query Match 3.4%; Score 113.5; DB 12; Length 2283;
Best Local Similarity 18.4%; Pred. No. 2.5;
Matches 145; Conservative 107; Mismatches 257; Indels 277; Gaps 36;

```

```

QY 22 VSANTVTVN-----GRNQRRRTGRQVSPDPNFTAAADLAQSLDANTVTPPAN 70
DB 266 VTANTITVKNCKMLKQYMTTSGNATYDQSTGIVTLQDAYS-----QKGAITLGR 315
QY 71 ISSMPEFNMAKIDLS-----DSIGWYFK----- 97
DB 316 IDSNSFPH--PSGKYNLANKYEGHNGDGDIGFAPSPGLVGTGLNGAAVIGLSNAP 373
QY 98 -----YLPDPAATBAPAVGYSKI-----PDGLVXSV 126
DB 374 FKLDTYHTSFKSNSAAKANADPSNVAAGARGAFYTTDSYGAATYTSSTADNAAKLV 433
QY 127 ---DARIREIYNEECVPTVDVSVPLDORQWLSLI-----PSFPM----- 162
DB 434 QPTNTPQDFPDIN--YNQDITVMT---VKYAGQVTRRISDMIAAGSTTNFSLSTASTAG 488

```

QY 163 -----FRT-AYVAVANVENKENSUDVNDLI--EMIANLADMRVYVDSQ----- 204
 Db 489 GATNLOOVQFGITEYESAVTQVRVYDVTTGKDIIIPKTVSGNVDDVVTIDNQSAALTRK 548
 QY 205 WINEFTN-DTTYVYRIRVRLRPVDPPEBGLVTVSPVRLTYKATICEANMPLTVQGF 263
 Db 549 GNYTSDVDSY-----ASTYN--DTNKTVKMTNMQGSVTTYFTVDKAPFTVVGQIT 599
 QY 264 IGGQVALPTPLSPQYDVSEAVYALHTLTPARPSSAALAFVWAGLPQ-----GTPAGT 317
 Db 600 VG--KTMNPIVLTTDNGTGTVTNTVT-----GLPSGLSYDSATNLSIGT 642
 QY 318 P-----AMEQASSGGYLTMRN--GTFPPAGSVSYVPEG-PALERTDP----- 358
 Db 643 PTKIGSTVTVVSTDOANNKSTTFTINVDYTAFT-----VTPIGDQSSSEVSPISPIK 697
 QY 359 -----NDGSWTDPASAGDVTFRQVAVDEVVYVNNPAGGSAPTF-----TVVPPSNAY 408
 Db 698 IATQDMSGN-----AVNTYVTLGSPGLTFDSTNTTISGTFPNIG 736
 QY 409 TNT-----VFRNTLLETSPSSRLRLPMPPADPQGTVANPKIEQ 448
 Db 737 TSTISVSTDASGNKTTTTFKYEYTRNSMSDSVTSSTQ-----SQSVSTSKXDSQ 789
 QY 449 SLKETLGCYLVH-----SKMRNPFV--QUTPASSFGAVSPNNPGYERTDLPDYGI 501
 Db 790 SASTSTGSIIVSTASTSTSTSVLSDSVASAKSLSTBSNSVSSSTSTSLVNSQSVSS 849
 QY 502 SFGQNNSTAVAHFRLSHSC-----SIVTKYQMGVNTNVTFRQGFAGLKNBEI 555
 Db 850 SMDSDASKSTSLSDSTSNSSSTKESLSTSDSLRTSTSLDSLSMSGSLSG-- 907
 QY 556 LCLADLRLTGVYPATDNFAAASFAANMLS--SVLKSATSIISKVBETAVGAA 612
 Db 908 -----SLSTISG--SSSTASLSLSTSNALSTSLSEASSTOSI-SISNLSANSQ 957
 QY 613 QSGLAK 618
 Db 958 SASTSK 963

RESULT 13
 US-09-738-626-6045
 ; Sequence 6045, Application US/09738626
 ; Publication No. US20020197605A1
 ; GENERAL INFORMATION:
 ; APPLICANT: NAKAGAWA, SATOSHI
 ; APPLICANT: MIZOGUCHI, HIROSHI
 ; APPLICANT: ANDO, SEIKO
 ; APPLICANT: HAYASHI, MIKIRO
 ; APPLICANT: OCHIAI, KEIRO
 ; APPLICANT: YOKOI, HARUHIKO
 ; APPLICANT: TATEISHI, MAKOTO
 ; APPLICANT: SENOH, AKIHITO
 ; APPLICANT: IKEDA, MASATO
 ; APPLICANT: OZAKI, AKIO
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-125
 ; CURRENT APPLICATION NUMBER: US/09/738, 626
 ; CURRENT FILING DATE: 2000-12-18
 ; PRIOR APPLICATION NUMBER: JP 99/377484
 ; PRIOR FILING DATE: 1999-12-16
 ; PRIOR APPLICATION NUMBER: JP 00/159162
 ; PRIOR FILING DATE: 2000-04-07
 ; PRIOR APPLICATION NUMBER: JP 00/280988
 ; PRIOR FILING DATE: 2000-08-03
 ; NUMBER OF SEQ ID NOS: 7059
 ; SOFTWARE: PatentIn ver. 3.0
 ; SEQ ID NO 6045
 ; LENGTH: 503
 ; TYPE: PRT
 ; ORGANISM: Corynebacterium glutamicum
 US-09-738-626-6045

Query Match 3.3%, Score 112.5, DB 10, Length 503,
 Best Local Similarity 19.9%, Pred. No. 0.27,
 Matches 114, Conservative 75, Mismatches 212, Indels 171, Gaps 26,
 QY 99 LBPAGTBSARAVERSKIPDGLVKEPSVABIRBYNECPVTVDVSPDLDRKSLGIF 158
 Db 51 LBPAGTBSARAVERSKIPDGLVKEPSVABIRBYNECPVTVDVSPDLDRKSLGIF 93
 QY 159 SPMFRTAVY-----AVANVENKENSUDVNDLIEMIANLADMRVYVDSQ 206
 Db 94 ---DDBTXYIFHLREGVLFNSGDPFRADSAKSIDRVK--TWNTNLSKSGMDVYSTXYI 148
 QY 207 NPTNDTTYVYRIRVRLRPVDPPEBGLVTVSDRLTYKATICEANMPLTVQGF 266
 Db 149 D-----DHTLKVSLVRF-----SNQWLW-----SNQTAIGAMWTEG 180
 QY 267 ---QVALPTPLSPQYDVSEAVYALHTLTPARPSSA--AALAFVWAGLPQGTAPAGTPAME 321
 Db 181 VDLATDPVGTGPY-----TYTHMAPGRALGFARADYWGQKPLDAATIRYFSDA 231
 QY 322 QASSGGYLTMRHNGTTFPAGSVSYV--LPSGFA-LERYDPNDG-----SWTDFAS 368
 Db 232 TAST-----NALSGDVYIWMAGAPQGLATLGTYTVBGTNGBMLSMNNGRA 281
 QY 369 AGDTVTFQV---AVDEVVYVNNPAGGSAPTFVYVPPSNAY--TNTV-----RNT 416
 Db 282 PPDVYVRQAVMFAIDRAVIDTALLEGYCTDVGVPVPTDWPYKSTJTPYDPDRARAL 341
 QY 417 LLETSPSSRLRLPMPPADPQGTVANPKIEQSLKETLGCYLVHSKMR-----NPFOJLT 472
 Db 342 LBSAGARGRITMSISLSPYQAASE-----ILISQLADVGFDPYIBST 385
 QY 473 PASBFGAVSPNNPGYERTDLPDYGIIRDSFGQNNSTAVAHFRLSHSCSIYTKYQME 532
 Db 386 ---EPFAYV-----LAQWGGQD-YWMSL--IAHYEPRIPTLFSPNYTLGPD 427
 QY 533 GVTNNTPPGQFRAHAGLKNBEILCLADLRLTGVYPATDN-FAAASFAANMLS 591
 Db 428 DT-----ETQALLAEADDSANBEVLMQOAVDRIMEQAVADNLMVANYIV 471
 QY 592 LKSEATSIISKVBETAVGAA 618
 Db 472 VMSPEITGIDPNVSGALSLIGKESGVAQ 503

RESULT 14
 US-10-245-802-18
 ; Sequence 18, Application US/10245802
 ; Publication No. US2003012413A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Edwards, John E.
 ; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITIONS AND METHODS TO VACCINATE AGAINST
 ; FILE REFERENCE: 013361.4003
 ; CURRENT APPLICATION NUMBER: US/10/245, 802
 ; CURRENT FILING DATE: 2002-09-13
 ; PRIOR APPLICATION NUMBER: US 09/715, 876
 ; PRIOR FILING DATE: 2000-11-18
 ; NUMBER OF SEQ ID NOS: 24
 ; SOFTWARE: PatentIn version 3.2
 ; SEQ ID NO 18
 ; LENGTH: 1443
 ; TYPE: PRT
 ; ORGANISM: Candida albicans
 US-10-245-802-18

Query Match 3.3%, Score 112.5, DB 15, Length 1443,
 Best Local Similarity 19.3%, Pred. No. 1.5, Indels 159, Gaps 29,
 Matches 118, Conservative 90, Mismatches 243,
 QY 49 FTAAQDLAOSLDANT---VTFPANISSMPFRNMAKXKIDLSDSISGMYFPYLDPA 104

```

Db 217 FSATMDV--TIDCSIHVGTNGLSNMP-----VSESFS-YTKTCTPS- 261
Qy 105 TESRAVEYSKIPGGLVYFVDABIR-----YNEBCPVTTDVSPIDG 150
Db 262 -----PILTYENVPAGYRP-IDSYKKSATATNGFNINYNIN--C-----MDG 304
Qy 151 RQMSLSIFPMFRAVAVANVENKEMSLDVVNDLIEMLNLIADMRVYVDSQMINF-- 208
Db 305 KKGNDDPLIF-----YTSYNSDAGSNGAAY-----VTRKVTDSITVITLTP 348
Qy 209 ---TNDTTYVVRIRVLRPTVDVDPTEGLVTVSDYRLTYKAITCEANMPTLVDOGFWIG 265
Db 349 FDPVTDKTKIEVIEPIPTTITTSYVGISTLSI--KATIGGATVVDV----- 398
Qy 266 GOYALTPISLPQYDVEAYALHTLTTPARPSAALAFWAGLPQGTAPAGTPANEQSS 325
Db 399 -PYHTTTITSIYTS--ATTSSTYNTPTSIDTVVQV-----PSPNPTVTTQFYS 448
Qy 326 GGYLTWRHNGTTFPAGSVSYVLPEGFALERVDPNDSMTDFASAGDTVPQGVAVDEVV 385
Db 449 GSVPT-TETVTGPGTDSVLIK-----EPHNPV-TTTEFYSSEFATT-----ETV 493
Qy 386 TNNPAGGSAPFTFVRVPSNAYTNTVFRNTLLETRPSSRLLEMPADFGQVANNPK 445
Db 494 TNNPEGTDV--IKEPHNPTVTTTEFYSSEFATT-----ETVNTPE 534
Qy 446 IEOSILKETLGCYLVHSMKRNPFQLTTPASSFGAVSFNNPGYERTRDLDPDTG--IDS 502
Db 535 GTDSVI-----VREPHNPVTTEFYSSEFATT--ETVNTPEGTDVIVREP 580
Qy 503 FDOWNSTAVAHFRSLSHSCSIVTKYQGMGVT-----NVNTPFQGFAGLKNBEI-- 555
Db 581 HNPVTTTEFYSSEFA-TTETITNYPGTDVIVREPHNPVTTEFYSSEFVTTETIT 639
Qy 556 -----LCIADDL-ATRLGVYPATDNFAAVSAFANMLSVLKSATSSIIKSYGE 606
Db 640 GPLGDSIVHDPLESSSSTAISSSDSNISSAQESSSVQSLTSDETSIVELSR 699
Qy 607 TAVGAQSGQL 616
Db 700 SDIPSSIGL 709

RESULT 15
US-10-187-253A-27
; Sequence 27, Application US/10187253A
; Publication No. US20030170612A1
; GENERAL INFORMATION:
; APPLICANT: Pichuanes, Sergio
; APPLICANT: Shyamala, Venkatakrishna
; TITLE OF INVENTION: DIAGNOSTIC ASSAYS FOR PARVOVIRUS B19
; FILE REFERENCE: CHIR-17194/03US / PPI7194.004
; CURRENT APPLICATION NUMBER: US/10/187,253A
; NUMBER OF SEQ ID NOS: 92
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 27
; LENGTH: 781
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: VP1 amino acid from
US-10-187-253A-27

```

Query Match 3.3%; Score 111.5; DB 12; Length 781;

Best Local Similarity 19.2%; Pred. No. 0.67;

Matches 126; Conservative 52; Mismatches 192; Indels 287; Gaps 29;

Qy 112 GEYSKIPDGLVYFVDABIR-----BIYNEBCPVTTDVSPVLDGRQMSLSIFSPMFRKAY 167

Db 85 GOLSDHPHALSSSSSHAEPRGEDAVLSSDLHKPGQVSVQLPG-----TNY 130

```

Qy 168 VAVANVENKEMSLDVVNDLIEMLNLIADMRV-----VYDSQMINFTNDT 212
Db 131 VQPN-----EIQAGPQSAVDASAARIDHDFRISQAKGINDPYTMTVADBEILKNIKET 186
Qy 213 TYVVRIRVLRPTVDVDPTEBGLVTVSDYRLTYKAITCEANMPTLVDOGFWIGQYALTP 272
Db 187 GFQAGV-----VQDY-FTLKG--AAFAVHPGQ-----SL 213
Qy 273 TSLPQYDVEAYALHTLTTPARPSAALAFWAGLPQGTAPAGTPANEQSSGGYLTWR 332
Db 214 PEVPAVNASERKP--SMTSVNSAEAS--TGAGGGGSPNVPYS--MMS- 254
Qy 333 HNGTTFPAGSVSYVLPSGFALERYP----- 358
Db 255 --GATFSANSVTCFSPROFLIP-IDPEHNYVFPSPASSCINAGSKAKVCTISPIWYS 311
Qy 359 -----NDGSMTDFASAGDTVPFRQVAVDEVVTTNPPAGG 393
Db 312 TPWRYLDPNALNLPFSPPLRFQHLIENTGS--IAPDALTVTISRIAVDV--TDKTOGG 365
Qy 394 -----SAPFTTVR--PSPNAYTNTVFRNTLLET 420
Db 366 VQVTDSTTGRLCMLVDHEKYRYYVGGQGDTLABELPIWVYFPQYAVLTGQDVNT-- 421
Qy 421 RPSRRLLEMPRPADFGQTVANNPKIEBSILKETLGCY-LVHSMKRNPFQLTTPASSFGA 479
Db 422 -----QGISGDSK--KLASSSAFVLHSS--FOLLGTGTAT 456
Qy 480 VSFN-----NPGYERTRDLDPDTGIRDSFDOWNSTAVAHFRSLSH 519
Db 457 MSYKRPVPPENLBSCSQHFYEMNPLXGSLGVPDTIG-----GDPKRSLLTH 505
Qy 520 -SCSIVTKYQGMGVTNNTVPPQGFAGLKNBEILCLADDLATRLGVYPA----- 572
Db 506 EDHAIQPFQFMFPGPLVNSVSTKGDSSGTGAKALTGISTGTSQWTRIS-LRPGFVSQPY 564
Qy 573 -----TDNFAAVSAFANMLSVLKSATSSIIKSYGTAAGA-----QSGIATLP 620
Db 565 HMDTDKVTGGINAI-----SHGQTYVGNABDKKEYQGVGRFP 602

```

Search completed: January 15, 2004, 17:23:18

Job time : 79.607 secs